Compaction Principle and Approaches to Biopolymer Structure Deciphering

I. P. Shabalkin and P. I. Shabalkin

Translated from *Kletochnye Tehnologii v Biologii i Medicine*, No. 2, pp. 109-115, April, 2010
Original article submitted May 29, 2009

On the basis of a principle of maximum compaction of biopolymer structure in a limited space derived from the concept about the existence of three different types of dimers in the studied structure, we propose a hypothesis according to which side groups in macromolecules (nitrogenous bases in nucleic acids or R-groups in proteins) are specifically distributed in different points depending on geometrical shape of dimer bending in the biopolymer chain backbone. The latter provides the possibility of complete deciphering of biopolymer structure.

Key Words: *compaction; biopolymer; deformation; dimer; structure deciphering*

Deciphering of biopolymer structure is important task in solving many problems in biology and medicine. A model of structural organization of biopolymer can be used for solving this task [10].

We previously proposed a two-dimensional model of shaping a nucleosome [11], a DNA chain unit, based on the principles of maximum compaction of DNA backbone determined by possible rearrangements of phosphate–sugar (P–S) or sugar–phosphate (S–P) monomers in the chain backbone. These rearrangements yield three types of dimers in DNA molecule: 1) (P–S)+(S–P), where the angle between the monomers is 60° [3,7]; 2) (S–P)+(P–S), where the angle between the monomers is 90° [2]; 3) (S–P)+(S–P), where angular deformation corresponds to 120° [18].

Let us consider the formation of maximally compact structure of the macronucleosomal loop of the first fragment of one primary DNA chain consisting of 144 monomers under conditions, when all (S–P)+(P–S) dimers of this fragment participate in its angular deformation. This fragment of DNA chain in two-dimensional space corresponds to the macronucleosomal loop (Fig. 1) with the following features:

1) symmetrical left and right parts of the loop. The existence of symmetrical parts allows transformation of the two-dimensional structure to three-dimensional one, *e.g.* by dividing the loop in the horizontal plane into 4 blocks (1-4) followed by bending of each block by 90° relative to the other one.

2) each block has side mini-loops; moreover, in blocks 1, 2, and 4 the symmetrical side mini-loops (a-i) can be bent at 90° relative to each other.

3) the macronucleosomic loop starts and ends with single-strand fragments of different lengths.

Generally, the three-dimensional structure of the nucleosome can be presented by a hollow rectangular structure with single-strand fragments at both ends (Fig. 2). This geometrical structure is a result of bending of each block (1-4) towards itself at an angle of 90°.

The role of mini-loops probably consists in stabilization of the structure through participation in compaction of the macro-loop during their cross-linking via a certain type of histone. We believe that selective binding of histones with a certain type of mini-loop is based on the principle of maximum compaction of nucleosome, therefore molecular weight of histones is important for the formation of the globular structure of nucleosome. Since histones [6] have different molecular weight (H4 11.3, H2B 13.8, H2A 14.0, and H3 15.3 kDa), they obviously occupy unequal area in the nucleosome. Hence, the nucleosome during its assembly should have sites near the core of the nucleosome with larger and smaller areas (for histones with

N. N. Blokhin Cancer Research Center, Russian Academy of Medical Sciences, Moscow, Russia. ***Address for correspondence:*** shabalkin@km.ru. I. P. Shabalkin

higher and lower molecular weights, respectively). The histones of different types most likely form a helix twisted along the existing grooves between the mini-loops. The histones probably are coiled around the macronucleosomic loop withoutside DNA strand. This assumption contradicts the hypothesis that DNA is wrapped around histones [16], but we believe that our scheme more precisely indicates the location of DNA mini-segments directly bound with certain type of histones in the nucleosomic loop.

Published data also indicate that the polypeptide is coiled around DNA chain [7].

The types of transformations described by us for the formation of nucleosomic moiety of one DNA chain are completely valid for the other DNA chain with consideration for rearrangement of monomers in its primary structure [11].

Bearing in mind that DNA molecule consists of two strands starting from either (S–P), or inverted (P–S) monomer [10], we assume that two types of nucleosomic moieties correspond them in different strands. These two moieties are the monomers similar to monomers of the primary DNA structure [10] and participate in the formation of the structure of each of two supramolecular DNA strands. However, it remains unknown where nucleosomes are arranged along the DNA chain according to certain regularity or this is a random process. If the primary structure of biopolymers obeys certain mathematical regularities, the assembly of supramolecular structural moieties corresponding to different levels of biopolymer organization should comply the regularities characteristic of its primary structure.

Let us assume that we assemble individual nucleosomes into a chain corresponding to a higher level of structural organization of DNA. To this end, it is reasonable to use the principle of primary structure based on permutation of (P–S) monomers in the chain backbone according to Fibonacci number sequence [10]. Let the monomer corresponding to the first member in the Fibonacci sequence (1^a) be a nucleosome starting from single-strand segment including 3 moieties belonging to moieties of the primary structure of DNA strand (Fig. 2), and let the monomer corresponding to the second member in the Fibonacci sequence (1^b) be a nucleosome starting from single-strand segment including 8 moieties also belonging to moieties forming the primary DNA structure. Thus, we divide the nucleosomes into two types by the number of monomers constituting the single fragments in the beginning and at the end of each nucleosome and thereby we use permutations allowing assembly of a chain composed, as was previously demonstrated [10] of a sum of Fibonacci numbers: $1^a+1^b+2+3+5+\ldots$*etc*. This chain can be presented as a "primary" structure

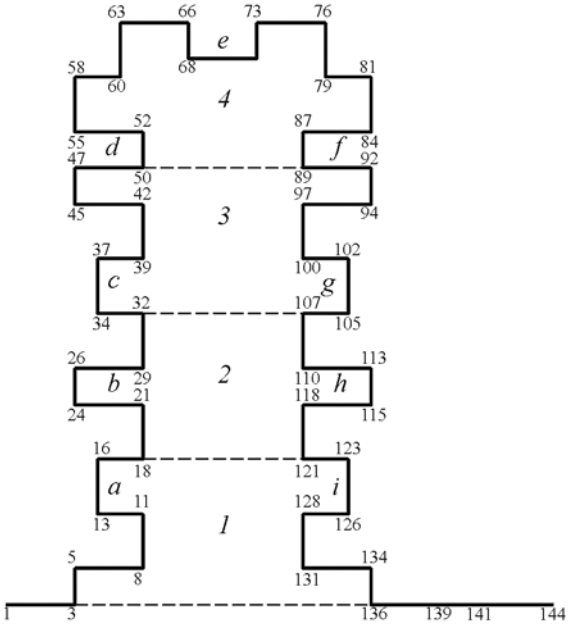

Fig. 1. Scheme of formation of macronucleosomal loop of DNA chain fragment under conditions when all [(S–P)+(P–S)] dimers of this fragment participate in its bending deformation. Dashed line divides the loop into blocks 1-4; a-i: mini-loops; 1-144: numbers of monomers.

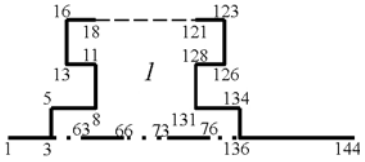

Fig. 2. General view of nucleosome loop package in a 3D structure in the form of one of four rectangular blocks with initial and terminal single-strand fragments. 1) block 1 (top view); 1-144: numbers of monomers; 1-3: initial single-strand fragment, 136-144: single-strand fragment terminating nucleosome assembly.

and transformed into a 3D structure according to the golden proportion principle [11].

Since supramolecular structure of DNA molecule is organized by the hierarchical principle, the formation of multicomplexes of nucleosomes should basically repeat the stages of assembly of a single nucleosome.

As a specific illustration, consider several stages of assembly of a single nucleosome.

According to published data, nucleosome is assembled in six stages [11]. During stage I, a small loop is formed consisting of two fragments: one of them includes 8 monomers and the other contains 5 monomers.

During stage II, a loop consisting of 73 monomers is formed. It includes 13 monomers forming a loop during stage I of nucleosome assembly and 60 additional monomers. During stage III, 63 monomers are added to the preexisting 73 monomers constitut-

ing the macro-loop structure. The assembly process during which new monomers are added to preexisting complex of monomers is completed at stage IV, when the last 8 monomers are added to 136 preexisting monomers.

Stages V and VI consist in the formation of the macronucleosome loop from the assembled structures belonging to two fragments of different length constituting the nucleosome.

In this model of macronucleosome loop we only demonstrated the role of [(S–P)+(P–S)] dimers in compaction of the loop, but did not take into account the peculiarities of compaction of the structure including other types of dimers [(S–P)+(S–P)] and [(P–S)+(S–P)]. The participation of the latter should considerably modify the structure of nucleosome loop, because of different strength of bonds between the monomers in different types of dimers [10]. Taking into account this fact and different angles between the monomers in these three types of dimers (60°, 90°, 120°) and bearing in mind that the process of structure compaction obeys the principle of least surface at the same volume we conclude that DNA should include geometrical forms that are presented by regular polygons. Moreover, DNA is a liquid crystal and combines the properties of a fluid and a solid, therefore it should include the elements of regular 3D crystalline lattice consisting of regular tetrahedrons and octahedrons [10,11]. We believe that these are the geometrical figures that participate in loop formation. The first mini-segment of the loop including four monomers consists of regular tetrahedron and octahedron, the moieties of regular crystalline lattice (Fig. 3), because crystalline lattice cannot be constricted with only one type of geometrical figure [1]. Interestingly, all three types of dimers are presented in this mini-segment.

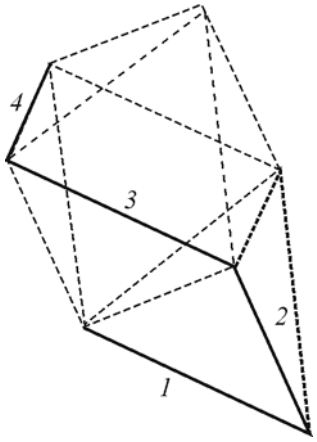

Fig. 3. Minimum crystalline lattice unit consisting of regular tetrahedron and octahedron. 1-4: fragment of "lagging" chain including monomers:

$$[\overset{1}{(P–S)}+\overset{2}{(S–P)}+\overset{3}{(S–P)}+\overset{4}{(P–S)}]$$

This model has some advantages over the structural DNA model proposed by Watson and Crick, because the existence of various bending deformations at the level of dimer formation considerably increases information capacity of the biopolymer structure under conditions of limited space.

Further analyzing our model we concluded that molecular structure of biopolymer is the limited space is based on compaction principle. According to this principle, complementarity of nitrogenous bases in DNA molecule is determined by interaction of a structure with high molecular weight and a structure with minimum molecular weight. The molecular weight of nitrogenous bases decreases in the following order: guanine→adenine→thymine→cytosine (151→135→136→123 Da, respectively), which reflects the known fact that guanine–cytosine and adenine–thymine form complimentary pairs in DNA molecule. Since complimentarity of nitrogenous bases and bending deformation of dimers in DNA molecule backbone are interrelated processes, a question arises about the nature of this phenomenon.

If the principle of maximum compaction of structures is valid at all levels of polymer organization, we admit that the dimers with an angle of 120° should contain guanine and adenine as the side groups, because these nitrogenous bases occupy maximum volume. Then cytosine should be present in dimers with bending deformation of 60° and thymine should be present in dimers bent at an angle of 90°, because this base occupies greater volume than cytosine due to the presence of CH_3-group.

Another question is whether we can determine which nitrogenous base corresponds to the first monomer in the dimer and which to the second one? The answer can be positive, if we take into account that the bending angle during dimer formation depends on the position of each second monomer relative to the first one. Let us consider a trimer consisting of monomers 1-3:

$$[\overset{1}{(P–S)}+\overset{2}{(S–P)}+\overset{3}{(S–P)}].$$

According to our hypothesis, angular deformation in dimer

$$[\overset{1}{(P–S)}+\overset{2}{(S–P)}]$$

is 60°, which implies binding of cytosine to monomer 2. The dimer

$$[\overset{2}{(S–P)}+\overset{3}{(S–P)}]$$

is bent at an angle of 120°. In this case, monomer 3 should contain purine (adenine or guanine), because

the most spacious nitrogenous bases can be more easily placed in dimers with bending angle of 120°.

In dimer [(S–P)+(P–S)] with bending angle of 90°, the second monomer should contain thymine as the side group. Further analysis of the distribution of nitrogenous bases in monomers of the DNA backbone depending on the role of the second monomer in each dimer leads us to a conclusion that the structure of biopolymer can be completely deciphered irrespective of undetermined position of nitrogenous bases: are they located outside or inside the chain.

There are three possible variants: 1) all nitrogenous bases are located inside the backbone consisting of two DNA strands. This variant corresponds to Watson–Crick model of DNA [19]. 2) Pauling and Corey model of DNA, according to which all nitrogenous bases are located outside the molecule backbone [17]. 3) The variant proposed by us and according to which some nitrogenous bases are located outside the polymer backbone.

Why we are not satisfied by other variants, in particular, the first one? First, this variant does not comply the hierarchical principle of macromolecular organization of DNA, *i.e.* the principle of base complimentarity is absent at high levels of compaction of such a giant polymer as DNA, because the bases hidden inside the lowest level of structural organization of DNA cannot form complimentary base pairs between the fragments belonging to higher supramolecular levels of structural organization, although the hierarchical principle implies maintenance (at least partial!) the base complimentarity rule at all levels of biopolymer organization. Second, the absence of bases outside the DNA molecule backbone makes impossible the regulation of growth, development, and functioning processes, because the surface of subunits belonging to different levels of structural organization carry no "signal molecules", *i.e.* "code" nitrogenous bases mediating activation or inactivation of various biochemical processes in the cell, because DNA carries genetic information about organism life.

The second DNA model does not meet the requirements for DNA structure for the following two reasons: 1) it does not include base complimentarity inside DNA chains, which impairs stability of the molecule; 2) DNA represents a structure consisting of three chains.

The third variant takes into account the advantages and disadvantages of the first two models and therefore we believe it corresponds to the optimal DNA structure. In this variant, the nitrogenous bases in each chain are distributed alternately: in each chain, one base in the dimer is located outside relative to the other base in the same dimer.

This distribution of bases in each chain is explained by the fact that during dimer formation the bending angle between the monomers does not allow placing both bases of the dimer in its internal space. Under these conditions, the position of sugar (pentose) in one monomer differs from that in the other monomer, similarly to the change in sugar position during pre-mRNA capping [5].

Figure 4 shows a fragment of DNA molecule including two DNA chains: left and right. Each chain consists of 4 monomers.

The backbone of the left chain includes monomers:

$$(\overset{1}{S–P})+(\overset{2}{P–S})+(\overset{3}{P–S})+(\overset{4}{S–P}),$$

and backbone of the right chain consists of monomers:

$$(\overset{1}{P–S})+(\overset{2}{S–P})+(\overset{3}{S–P})+(\overset{4}{P–S}).$$

The left chain is "leading", because is begins from the most ancient nucleotide, deoxyadenylic acid [12]. The right "lagging" chain is shifted by one nucleotide relative to the left chain. Due to this, the "lagging" chain can form the first complimentary pair of DNA molecule: thymine+adenine. This became possible, because the nitrogenous base of the second monomer in [(S–P)+(P–S)] dimer is determined in the first dimer of the "leading" chain.

In the first monomer of the "leading" chain, the nitrogenous base adenine should be located outside the chain, because it performs the function of the first signal molecule, the start point for the formation of DNA structure. The scheme of partial location of nitrogenous bases outside of chain backbones stabilizes (via the formation of complimentary pairs) the fragments of supramolecular structures at different levels of molecular organization of DNA. This suggests that the model of structural organization of DNA proposed by us better (compared to the first two models) corresponds to the role of DNA in organism life.

In is interesting to consider the mechanism of functioning of DNA polymerases, because these processes include break and transfer of DNA chain to another active center. It is known that the chain can be broken due to interaction of the growing active center with introduced second monomer increasing chain rigidity. Chain transfer is a transfer of active center of the molecule to another molecule initiating the growth of a new chain [14]. These processes directly depends on [(S–P)+(P–S)] dimers. The point is that enzymes catalyzing the growth of polynucleotide chain (*e.g.* DNA polymerase) do not work in this case, because they normally attach the next monomer only to the 3'-hydroxyl group of the sugar in the end

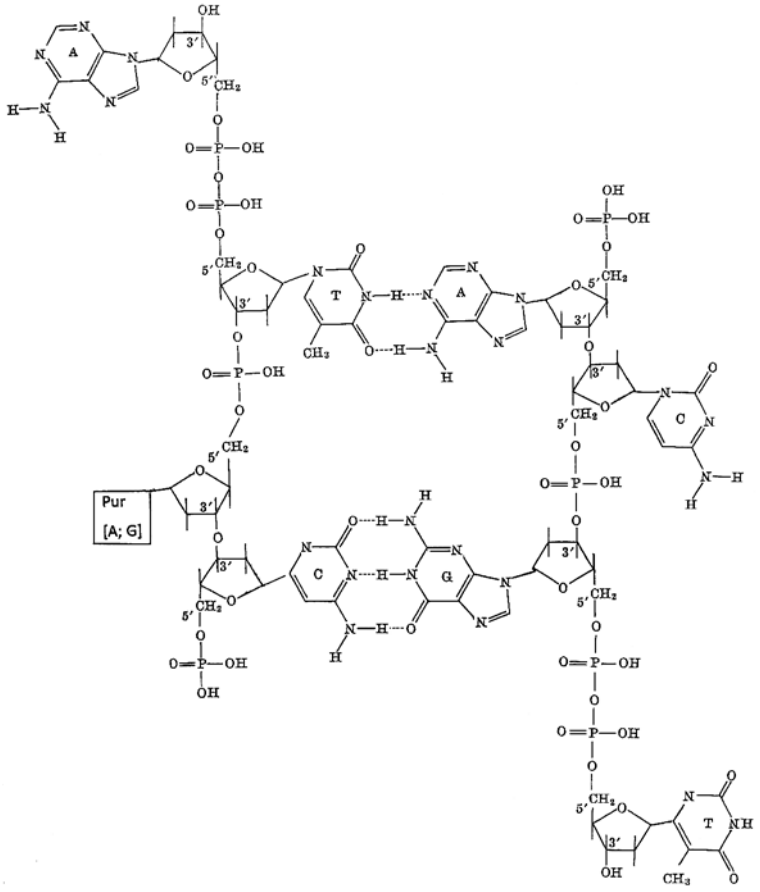

Fig. 4. Fragment of double-strand DNA: some nitrogenous bases in each chain of DNA molecule are located outside the polymer backbone. On the left: "leading" chain, on the right: "lagging" chain.

monomer, which should not be phosphorylated at the stage of chain elongation. When DNA polymerase approach the dimer [(S–P)+(P–S)], the chain is broken. In order to continue the process of chain elongation, the enzyme should find the transition point.

Let us discuss these processes in more detail on the basis of matrix (template) backbone monomer synthesis of the "leading" and "lagging" chains also presented by their monomers (Fig. 5). Eukaryotes have several types of DNA polymerases [5]; let the first type of DNA polymerase start working and transcribe the "leading" chain, where the first monomer is (S–P). After synthesis of the first daughter domain (P–S) complimentary to the first monomer in the "leading" chain matrix, transcription of this chain is terminated, because the second monomer in the chain starts from the phosphate group. The synthesis of the daughter DNA chain is paused (!) for a while. Several processes take place in the cell during this pause. Break hydrogen bond between the nitrogenous bases of the second monomer of the "leading" chain and the first monomer of the "lagging" chain leads to outside displacement of the sugar and nitrogenous base (adenine) of the first monomer in the "lagging" chain. This phenomenon was observed by investigators using a magnetic forceps for untwisting of the B-DNA [15].

The first nucleotide of the "lagging" chain in the new position becomes available to other DNA polymerase, which can use single-strand structures starting from (P–S) monomers as the template. After synthesis of the first daughter monomer (S–P) in the "lagging" chain, this DNA polymerase cannot work on the same template, because monomers are unavailable, because they are shield by type I DNA polymerase. When type I DNA polymerase finds a new matrix, in particular the second monomer of the "lagging" chain and moves to this new active center, type II DNA polymerase jumps from the "lagging" chain to the "leading" one and starts DNA synthesis using the second and third monomers of this chain as the template. Generally, during replication the synthesis of the daughter fragment on one template DNA chain is broken and the catalyzing enzyme moves to a new active center, *i.e.* to the second template DNA chain. This is a cyclic process associated with changes of active center. During replication, the "leading" chain is by one nucleotide

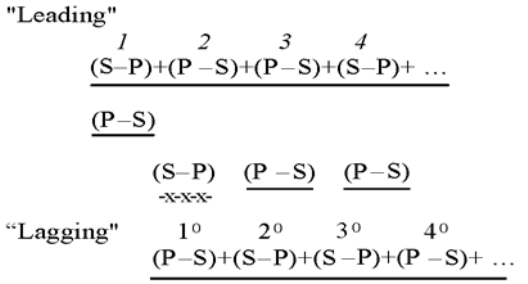

Fig. 5. Scheme of functioning of DNA polymerases involving break and transfer of DNA chain to another active center.

1
(S–P): the first monomer of the "leading" DNA chain ("leading");

1
(P–S): the first monomer of the "lagging" DNA chain ("lagging").

ahead to the "lagging" chain. Every time the pause occurs between termination and activation of DNA synthesis in new active center, when no DNA is synthesized; hence, DNA synthesis has a successively interrupted nature.

It should be noted that the hypothesis on the interrupted nature of DNA synthesis was put forward by some authorities [4,7,9,13], and was even experimentally proven [8]. Thus, we can conclude that the process of biochemical evolution of biopolymers led to the appearance of mechanisms based on the principle of maximum structural compaction in a limited space. This was associated with selective distribution of nitrogenous bases or R-groups of proteins in biopolymer macromolecules to certain locations determined by bending deformations of dimers in the biopolymer backbone.

The DNA model developed by us has some advantages over the existing Watson and Crick DNA model and Pauling protein model.

First, in both these models the monomers in polymer chain backbones strictly follow each other in the sequence preset by the model. This imposes restrictions to polymer structure: the molecule cannot be most compactly packed in a limited space, because this structure does not comply the requirements of optimal bending deformations in polymer chain. In other words, the processes related to monomer permutations in polymer backbone were not taken into account despite the fact that these monomer permutations yield mini-segments with different bond strength between the monomers in the chain. Different nature of bonds between the monomers attests to the existence of relatively nonrigid segments increasing chain flexibility, which provides additional possibilities for spatial compaction of the molecule and for coordination binding of ligands.

Monomer permutation in the polymer chain leads to creation of absolutely new molecular biosystems meeting the requirements to maximal compaction of the molecule in a limited space and increasing information capacity of biopolymer structure due to the existence of numerous variants of monomer compaction. It should be noted that these processes are not random, but obey mathematical laws. Based on the belief that polymerization of the biopolymer chain naturally proceeds in accordance to mathematical regulation known as Fibonacci sequence, we determined the course of formation of the backbone of primary polymer structure consisting of monomers including two different chemical groups, which provides the possibility of constructing molecular system with predicted spatial arrangement of certain fragments of the chain. The submolecular fragments are formed according to another mathematical regularity, golden proportion. According to the golden proportion law, the calculated minimum fragment participating in the assembly of DNA chain nucleosome should consist of at least two mini-segments, the larger containing 8 monomers and the smaller consisting of 5 monomers. The geometry of the fragment within the proposed new model meets the requirement of minimum surface at the same volume, which implies the existence of a structural unit composed of regular tetrahedron and octahedron. These structures act as information condensers and transmitters. This structure corresponds to chain fragments where acoustic signals propagate without energy loss. The fact that this system consists of structural elements presented by regular geometric figures creates prerequisites for assignment of biology and medicine to exact sciences at the modern level of nanotechnology development. For instance, deciphering of information encoded by biopolymers provides the possibility for the synthesis of drugs strictly identical by their stereo structure to biological targets in cells and for the creation of artificial compounds regulating functional activity of biopolymer.

REFERENCES

1. T. Varga, *Mathematics* [in Russian], Moscow (1978).
2. D. V. Veizer and A. K. Kleiton, *Inorganic Polymers* [in Russian], Moscow (1965), pp. 31-83.
3. W. Zenger, *Principles of Structural Organization of Nucleic Acids* [Russian translation], Moscow (1987).
4. A. Cornberg, *Synthesis of DNA* [Russian translation] Moscow (1977).
5. L. I. Patrushev, *Gene Expression* [in Russian], Moscow (2000).
6. M. Singer and P. Berg, *Genes and Genomes* [Russian translation], Moscow (1980).
7. J. Finean, *Biological Ultrastructure* [Russian translation], Moscow (1970).
8. I. P. Shabalkin, *Arkh. Anat*., **71**, No. 9, 29-35 (1976).
9. I. P. Shabalkin, *Thitologiya*, **19**, No. 5, 474-477 (1977).
10. I. P. Shabalkin and P. I. Shabalkin, *Kletoch. Tekhnol. Biol.*

Med., No. 4, 211-215 (2008).
11. I. P. Shabalkin and P. I. Shabalkin, *Ibid.*, No. 4, pp. 214-218 (2009).
12. I. P. Shabalkin, P. I. Shabalkin, and A. S. Yagubov, *Zh*. *Evoluts. Biokhim. Fiziol.*, **49**, No. 5, 488-495 (2003).
13. M. Eigen and P. Schuster, *The Hypercycle: A Principle of Natural Self-Organization* [Russian translation], Moscow (1982).
14. *Encyclopedia of Polymers* [in Russian], Moscow (1977), Vol. 2, pp. 383-384, 405-406, 959.
15. J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, *Proc. Natl. Acad. Sci. USA*., **95**, No. 24, 14152-14157 (1998).
16. A. Mirzabekov, *Quarterly Review of Biophysics*, **13**, 255-295 (1980).
17. L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. USA*., **39**, 84-97 (1953).
18. I. Smith, M. J. Smith, and C. F. Dore, *Chromatographic and Electrophoretic Techniques,* **2**, Ch. 16, 409-453 (1976).
19. J. D. Watson and F. H. C. Crick, *Nature*, **171**, 737-738 (1953).