

ELASTIC ROD MODEL OF RNA 3D STRUCTURE

E.E.Kozyreva

State Research Institute of Genetics and Selection of Industrial Microorganisms, 1st Dorozhny proezd
1, 113545 Moscow

E.I.Kugushev, E.L.Starostin*

M.V.Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Miusskaya pl. 4,
125047 Moscow

(A draft version, 12.07.01)

INTRODUCTION

The spatial shape of biological molecules is known to be one of the important determinants of their biochemical properties. Therefore, the specification and prediction of three-dimensional folding of biological macromolecules is one of the most challenging and fundamental problems of molecular biology. This work is devoted to the search of the proper technique for solution of this problem for an important class of biopolymers: ribonucleic acids. Three-dimensional structure of large RNAs is currently understood not so well as that of the other biological macromolecules [8], though one can observe a significant progress in this area [1].

A problem of prediction of approximate large-scale 3D structure of an RNA molecule from its secondary structure is considered. Both a mathematical model and its computer implementation are presented. An RNA molecule is treated as a system of linked basic structural elements (stems and single-stranded fragments including loops of various types) modelled by elastic rods. A numerical procedure is developed for computation of shapes of the RNA elements and for assembling the whole molecule.

MODEL DESCRIPTION

An approach is proposed for investigation and analysis of RNA spatial shape. It is based on the theoretical methods well adapted to prediction of the DNA large-scale structure [6,7,12]. The tertiary structure of RNA molecule is built from basic structural elements with pre-computed 3D shape. These basic elements are stems, loops of various types and other single-stranded fragments.

Every loop is modeled as a closed contour consisting of a number of thin curvilinear elastic rods linked at their ends by absolutely rigid cross-bonds simulating Watson-Crick interactions (we assume that the cross-bond joins the central points of the paired nucleotides). The number of the rods is equal to the number of the branches of the loop. With adequate choice of the elastic and geometrical parameters of the model rod, its shape approximates the large-scale 3D structure of the structural element of the RNA molecule. It would appear reasonable that, when unstressed, all the rods constitute the helical structure of a strand of the RNA in A-form.

We consider (i) dangling ends, (ii) single-stranded fragments that only join two stems and (iii) double-stranded parts (stems) as fixed curved and twisted rods which are in the unstressed state and which are represented by single ((i) and (ii)) or double (iii) regular helices. The other single-stranded parts of the molecule (loops of various type) are treated as stressed. A spatial equilibrium shape of a loop is determined by finding the solution of the system of the boundary value problems (BVP) corresponding to the rod fragments, which satisfy the geometrical constraints at their ends. The application of the continuous elastic rod model to the single-stranded fragments might be justified by a consideration that the elasticity can effectively mimic the actual properties of the chain of nucleotides arising due to base stacking interactions [10].

Let us consider a single-stranded fragment of a loop, which is modeled as a thin elastic rod of length L . The rod is assumed to be inextensible and unsharable. The vector function $\bar{r}(s)$, $s, 0 \leq s \leq L$, describes the centreline of the rod and s is the arclength. Denote by $\bar{e}_i, i = 1, 2, 3$, the unit vectors of the principal axes of the strain tensor, where \bar{e}_1 is the tangent vector and

*Current address: Département de Mathématiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

$$\frac{d\bar{r}}{ds} = \bar{e}_1 \quad (1)$$

because of the inextensibility assumption.

Let B_i are the corresponding elastic coefficients (which are taken to be constant along the rod), then B_1 is the torsional stiffness, and B_2 and B_3 are bending stiffnesses. M_i stand for the components of the moment $\bar{M}(s)$; w_i and w_i^0 are the components of the Darboux vectors \bar{W} and \bar{W}^0 for the stressed and unstressed rod, respectively. We apply the generalized Hooke law as the constitutive relation

$$M_i = B_i (w_i - w_i^0), \quad i = 1, 2, 3 \quad (2)$$

The equilibrium of the rod fragment is governed by the following system of equations [15]:

$$\begin{aligned} \frac{d}{ds} (B_1 (w_1 - w_1^0)) + B_3 w_2 (w_3 - w_3^0) - B_2 w_3 (w_2 - w_2^0) &= 0 \\ \frac{d}{ds} (B_2 (w_2 - w_2^0)) + B_1 w_3 (w_1 - w_1^0) - B_3 w_1 (w_3 - w_3^0) - F_3 &= 0 \\ \frac{d}{ds} (B_3 (w_3 - w_3^0)) + B_2 w_1 (w_2 - w_2^0) - B_1 w_2 (w_1 - w_1^0) + F_2 &= 0 \end{aligned} \quad (3)$$

and

$$\begin{aligned} \frac{d}{ds} F_1 + w_2 F_3 - w_3 F_2 &= 0 \\ \frac{d}{ds} F_2 + w_3 F_1 - w_1 F_3 &= 0 \\ \frac{d}{ds} F_3 + w_1 F_2 - w_2 F_1 &= 0, \end{aligned} \quad (4)$$

where $\bar{F} = (F_1, F_2, F_3)$ is an external force applied to the ends of the rod fragment (no other external forces and moments are taken into account). These equations include six parameters: B_i , w_i^0 , $i = 1, 2, 3$. The equations are closed with respect to variables w_i , F_i , $i = 1, 2, 3$, but knowing their values is not sufficient for the computation of the rod shape in the 3D space.

To close the system of the equations, we add an equation expressing the fact that the vectors \bar{e}_i , $i = 1, 2, 3$, are constant in the principal axes of the strain tensor: $\frac{d\bar{e}_i}{ds} = \bar{W} \times \bar{e}_i$, $i = 1, 2, 3$.

Since $\bar{e}_2 = \bar{e}_3 \times \bar{e}_1$, two equations for $i = 1, 2$ are enough. Applying the decomposition $\bar{W}(s) = w_1 \bar{e}_1 + w_2 \bar{e}_2 + w_3 \bar{e}_3$, we have:

$$\frac{d\bar{e}_1}{ds} = w_3 \bar{e}_3 \times \bar{e}_1 - w_2 \bar{e}_3, \quad \frac{d\bar{e}_3}{ds} = w_1 \bar{e}_1 \times \bar{e}_3 + w_2 \bar{e}_1 \quad (5)$$

We should notice that these equations are redundant as the orientation of the trihedral can be expressed in three parameters instead of six (\bar{e}_1 and \bar{e}_3).

Search for the solution of the system (3-5) is equivalent to the minimization of the integral elastic energy of the rod with known end constraints. It is not necessary to integrate the equations (4) since the force \bar{F} is constant in the laboratory reference frame.

Thus, the system of 12 equations (3, 5, 1) is closed and it can be integrated if the initial values are given:

$$\vec{r}(0) = \vec{r}_0, \quad \vec{e}_1(0) = \vec{e}_{10}, \quad \vec{e}_3(0) = \vec{e}_{30}, \quad (6)$$

$$\mathbf{w}_i(0) = \mathbf{w}_{i0}, \quad i = 1, 2, 3, \quad \vec{F} = \vec{F}_0, \quad (7)$$

The first three conditions (6) determine the position and orientation at the initial point (the 3' end as applied to the RNA) of the rod. The fourth and the fifth conditions (7) define the moment and the force acting on the initial point of the rod.

To compute the spatial shape of RNA, it is necessary to find a configuration of the rod under the conditions on its ends. The position and orientation of both ends are defined while the forces and the moments at the initial point are determined as a result of the solution of BVP. More precisely, the boundary conditions are:

$$\vec{r}(0), \vec{r}(L), \vec{e}_1(0), \vec{e}_1(L), \vec{e}_3(0), \vec{e}_3(L) \quad (8)$$

and the solution of the BVP means the correct choice of the quantities:

$$\mathbf{w}_i(0), \quad i = 1, 2, 3, \quad \vec{F}$$

Now we illustrate the formulation of the boundary conditions on an arbitrary hairpin loop.

The thin elastic rod modeling the hairpin loop is joined smoothly to the both strands of the double helix region. Therefore the position and orientation of the principal trihedral at the initial (final) point of the rod coincide to those at the final (initial) point of the first (second) strand of the double helical fragment. The parameters of the double-stranded section are known and correspond to the parameters of the A-form of RNA [16].

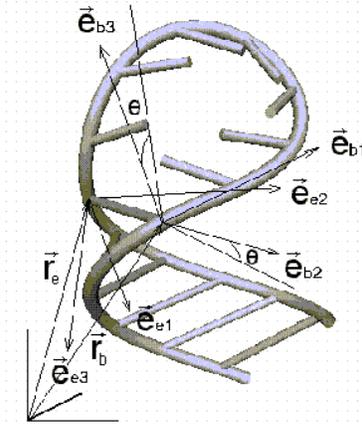


Fig. 1. Boundary value constraints for the hairpin loop.

$$\vec{r}(0) = \vec{r}_b, \quad \vec{r}(L) = \vec{r}_e, \quad \vec{e}_1(0) = \vec{e}_{b1}, \quad \vec{e}_1(L) = \vec{e}_{e1}, \quad \vec{e}_3(0) = \vec{e}_{b3}, \quad \vec{e}_3(L) = \vec{e}_{e3}$$

Fig. 1 demonstrates the description of the BVP for the hairpin loop. The angle \mathbf{q} is a rotation angle of the principal axes of the strain tensor with respect to the Frenet trihedral for the non-stressed rod.

To compute a spatial configuration of a multi-branched loop we have to find the equilibrium shape of a closed contour consisting of a number of single-stranded fragments modelled by elastic rods. Every pair of consecutive rods is joined by a rigid cross-bond. Self-interactions of remote parts of the loop as well as of the whole molecule (tertiary interactions) are not taken into account.

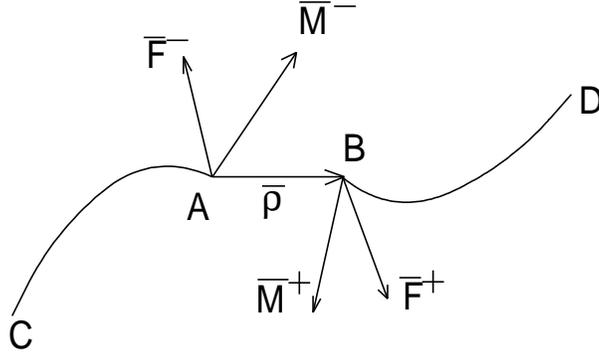


Fig. 2. Two thin elastic rods CA and BD joined by a rigid cross-bond AB.

Consider a pair of neighbour fragments of the thin elastic rods CA and BD of lengths L_1 and L_2 , respectively, and let the cross-bond AB join them. The points A and B are the ends of the double helical region that contains a cross-bond modeling the Watson-Crick bond. Let us assume for the moment that we have found the initial conditions and integrated the equilibrium equations from point C to point A, i.e. from 0 to L_1 . Denote the force $\bar{F}(L_1) = \bar{F}^-$ and the moment $\bar{M}(L_1) = \bar{M}^-$ applied to the end of the fragment (point A). Therefore, the force $-\bar{F}^-$ and moment $-\bar{M}^-$ are applied to the rigid cross-bond in the point A. Also designate $\mathbf{w}_i(L_1) = \mathbf{w}_i^-$, $\bar{e}_i(L_1) = \bar{e}_i^-$, $i = 1, 2, 3$. These quantities are specified by the integration. We have

$$\bar{M}^- = B_1(\mathbf{w}_1^- - \mathbf{w}_1^0)\bar{e}_1^- + B_2(\mathbf{w}_2^- - \mathbf{w}_2^0)\bar{e}_2^- + B_3(\mathbf{w}_3^- - \mathbf{w}_3^0)\bar{e}_3^- \quad (9)$$

Let $\bar{r}^- = \bar{r}(L_1)$ be a position of point A in the absolute space. To calculate the shape of the second elastic rod fragment, we need to know the initial values:

$$\bar{r}(0) = \bar{r}^+, \bar{e}_1(0) = \bar{e}_1^+, \bar{e}_3(0) = \bar{e}_3^+, \mathbf{w}_i(0) = \mathbf{w}_i^+, i = 1, 2, 3, \bar{F}(0) = \bar{F}^+ \quad (10)$$

(The second fragment BD is parameterized from 0 to L_2 .)

The position and orientation of the cross-bond AB can be approximated on the basis of the following constraints:

- the cross-bond is inclined at the angle of 70° to the symmetry axis of the double helix;
- the length of the cross bond is equal to 19 \AA that corresponds to the diameter of the double helix in the B-form.

We assume that the single-stranded fragments CA and BD are joined smoothly to the strands of the double helical region, therefore, the orientation of the principal trihedral of the second rod fragment in the initial point B can be derived from the helix structure of the A form.

Then

$$\bar{e}_i^+ = G\bar{e}_i^-, \quad i = 1, 2, 3, \quad (11)$$

where G is a known orthogonal matrix. The position of the cross-bond $\bar{\mathbf{r}} = \overline{AB}$ in the axes of the principal trihedral can be represented by

$$\bar{\mathbf{r}} = \mathbf{r}_1\bar{e}_1^- + \mathbf{r}_2\bar{e}_2^- + \mathbf{r}_3\bar{e}_3^- \quad (12)$$

and the initial position of the second fragment is

$$\bar{r}^+ = \bar{r}^- + \bar{\mathbf{r}} \quad (13)$$

In our model the components \mathbf{r}_i as well as the matrix G are calculated from the geometry of the A-form.

To complete the initial conditions for the second fragment, it is necessary to find \mathbf{w}_i^+ , $i = 1, 2, 3$, and $\bar{\mathbf{F}}^+$. Actually we should find the force $\bar{\mathbf{F}}^+$ and the moment $\bar{\mathbf{M}}^+$, applied in the point B of the rigid cross-bond. To write out the equilibrium equations for the cross-bond, it is essential to equate all the forces and moments in the point B to zero. For the forces we have:

$$\begin{aligned} -\bar{\mathbf{F}}^- + \bar{\mathbf{F}}^+ &= 0, \\ \text{or} \\ \bar{\mathbf{F}}^+ &= \bar{\mathbf{F}}^- \end{aligned} \quad (14)$$

The moments with respect to the point B include the moments in the points A ($-\bar{\mathbf{M}}^-$) and B ($\bar{\mathbf{M}}^+$) as well as the moment of the force in the point A relative to point B ($(-\bar{\mathbf{r}}) \times (-\bar{\mathbf{F}}^-)$). If we equate the sum of these vectors to zero we shall get the second equilibrium equation for the cross-bond:

$$-\bar{\mathbf{M}}^- + \bar{\mathbf{M}}^+ + \bar{\mathbf{r}} \times \bar{\mathbf{F}}^- = 0 \quad \text{or} \quad \bar{\mathbf{M}}^+ = \bar{\mathbf{M}}^- - \bar{\mathbf{r}} \times \bar{\mathbf{F}}^- \quad (15)$$

We know that

$$\bar{\mathbf{M}}^+ = B_1^+(\mathbf{w}_1^+ - \mathbf{w}_i^0)\bar{\mathbf{e}}_1^+ + B_2^+(\mathbf{w}_2^+ - \mathbf{w}_2^0)\bar{\mathbf{e}}_2^+ + B_3^+(\mathbf{w}_3^+ - \mathbf{w}_3^0)\bar{\mathbf{e}}_3^+$$

and hence

$$\mathbf{w}_i^+ = \frac{(\bar{\mathbf{M}}^+, \bar{\mathbf{e}}_i^+)}{B_i^+} + \mathbf{w}_i^0 \quad (16)$$

Therefore the expressions (14-16) allow us to find the initial data for the equilibrium equations of the second fragment.

As the multi-branched loop is described as a set of several elastic rod fragments joined by rigid cross-bonds modeling Watson-Crick interactions, it is now straightforward to write out a BVP for the equilibrium shape of the loop in the same way as it was stated for the hairpin loop.

Let the multi-branched loop have n single-stranded fragments with the lengths L_i . To compute its 3D shape we should find the shape of the rod chain with known position and orientation at the initial point (3' end) of the first fragment and at the terminal (5' end) point of the last fragment though the forces and the moments at the initial point are unknown. Both forces and moments are supposed to be calculated in the process of the BVP solution. More precisely, the boundary conditions are presented by (10).

The BVP above is solved numerically by the shooting method. Its solution depends on the elastic coefficients of the model and on the geometric properties of the unstressed state.

To sum up, the shape of a loop is determined by its boundary conditions, namely, by the position and the orientation of the principal trihedral in the initial (3') and terminal (5') points of the strands of the double-stranded fragment of the A-form of RNA in that place where the last Watson-Crick bond passes before the loop. These constraints are applied to the corresponding ends of the rods modelling the loop. Also, we accept a simplification that the Watson-Crick bond may be represented as a rigid constraint connecting the central points of the complementary nucleotides.

COMPUTATION OF FINAL 3D SHAPE

The procedure of computation of 3D structure has two stages.

At the first stage the 3D shapes of basic elements (stems, loops, dangling ends) should be calculated. Only elements that are present in the secondary structure of given RNA are processed. The shape of any loop is a result of the solution of the corresponding BVP. Stems and dangling ends are fixed. The stem is treated as a pair of interwound cross-bound rods in a relaxed state while the loop's rods are in the stressed state.

At the second stage the whole structure of the molecule is assembled by means of successive addition of basic elements. Since the boundary conditions correspond to A-form and the double helices of stems are also in A-form, the elements are glued smoothly that means that all stem rods and loop rods constitute one rod with continuous tangent. The ends of this composite rod correspond to the 3' and 5' ends of the molecule. As it has been already mentioned above, the cross-bonds are treated as

absolutely rigid and their position and orientation with respect to the rod are fixed. Therefore, the shape of the composite rod is completely identical to the shape which is taken on by one continuous rod of the same length that gives a solution of the BVP. The latter involves all the constraints that arise due to the cross-bonds.

RESULTS

We applied the above procedure to two different types of RNA. The numerical values of the parameters were chosen as follows.

In the simplest case we can try $B_1 = B_2 = B_3$ that corresponds to the ...

$$\mathbf{w}_2 = 0.0, \mathbf{w}_3 = 7.549074e-002, \mathbf{w}_1 = 2.942040e-002 \text{ [rad/\AA]}$$

In particular, we computed the structures of the Yeast Phenylalanine Transfer RNA (Fig. 3a, 3b). Its tertiary structure was determined by X-ray analysis (Fig. 3c) and described in detail in [14]. In the figures the view direction is chosen approximately perpendicular to the “plane” of the molecule.

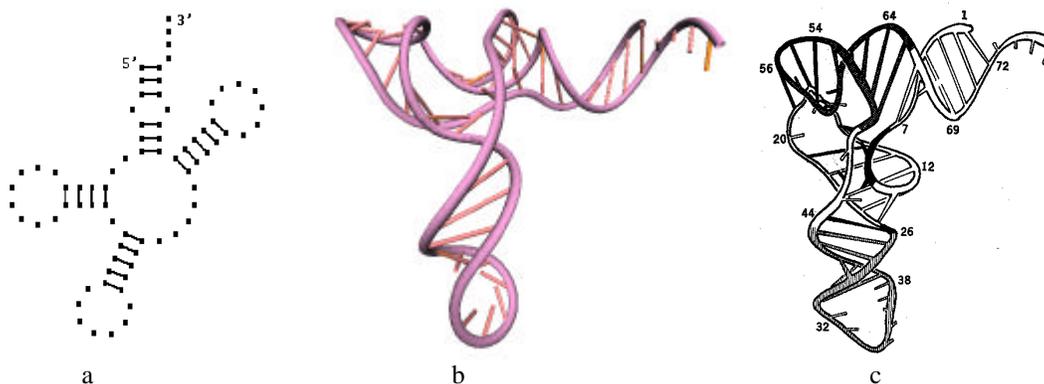


Fig. 3. Yeast Phenylalanine Transfer RNA.

The comparison of this RNA with the computed shape shows that even such a simple model allows one to get some qualitative resemblance of the overall conformation. Namely, the result of modelling catches the following experimentally confirmed features of the molecule’s conformation:

- the molecule as a whole is somewhat flattened;
- it has an L-shape conformation;
- the acceptor stem is at an approximately right angle to the anticodon stem;
- the two other stems are in the position that facilitates the tertiary interactions between nucleotides of their hairpin loops.

The secondary structures of RNAs in Figs. 4-6 are taken from [9,11].

It should be noted that these features are more or less common to short transfer RNAs (Fig. 4-6).

A ribosomal RNA is shown in the last Fig. 7.



Fig. 4. tRNA: Arginine Cenorhabdi. Elg.

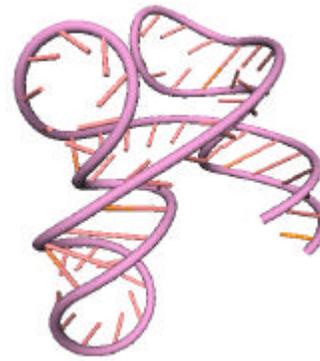
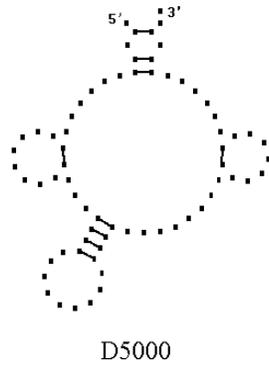


Fig. 5. tRNA: Asparagic Acid Asterina Pectini.

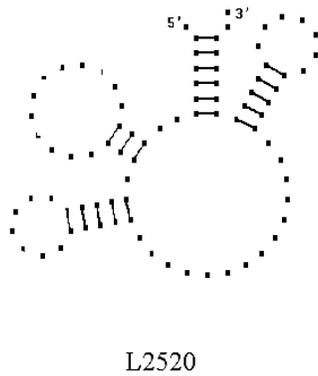


Fig.6. tRNA: Leucine Eugelna Gracilis.

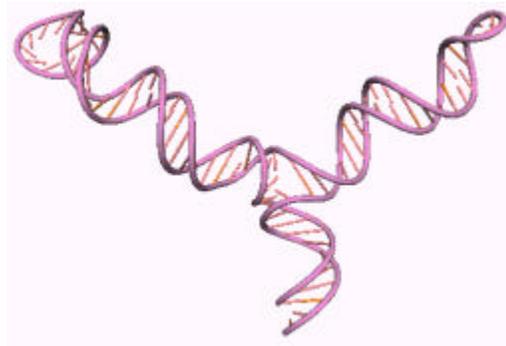
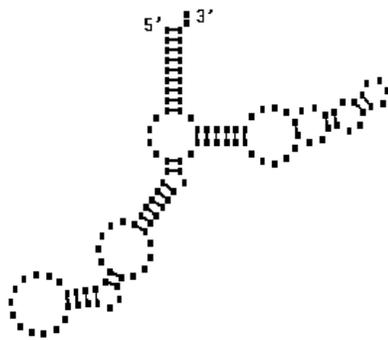


Fig.7. 5S rRNA of human.

CONCLUSION

The elastic rod model can be applied to prediction of an approximate 3D shape of not only DNAs but RNAs, as well.

It should be noted that the secondary structure may be affected by tertiary interactions [5]. In this respect, a means for the fast computation of a rough three-dimension configuration may be possibly used in the iterative procedure for the search of the optimal secondary and tertiary structures. It is our belief, that the approach described may eventually provide such a means. Besides, the presented elastic rod model may serve as an initial approximation for a more elaborated procedure of the shape computation at the base (or even atomic) level.

The significant advantage of the model proposed is a small number of the parameters defining the structure of the molecule. At the same time it is a priori clear that this model may produce only large-scale approximation of real polynucleotide chains because, among other things, it does not take into account effects of tertiary interactions between distantly located bases, in particular, in fixed stems

and dangling ends. Now we are working on further development of the model by taking into account the heterogeneity of nucleotide chains and on verification of the results against input data and testing the robustness of the model relative to uncertainty of the parameter values.

ACKNOWLEDGMENT

This work was supported by the Russian Foundation for Basic Research under grant N 99-01-00029.

We especially thank Prof. S.V. Mashko from the State Research Institute of Genetics and Selection of Industrial Microorganisms for his invaluable help with the model improvement and for useful discussions.

REFERENCES

1. Ban N., Nissen P., Hansen J., Moore P.B., and Steitz T.A., 2000, The complete atomic structure of the large ribosomal subunit at 2.4 Å Resolution, *Science*, 289, 5481, pp. 905-920.
2. Burkard M.E., Kierzek R., Turner, D.H., 1999, Thermodynamics of unpaired terminal nucleotides on short RNA helices correlates with stacking at helix termini in larger RNA, *Journal of Molecular Biology*, 290, pp.967-982.
3. Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H., 1999, Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure, *Journal of Molecular Biology*, 288, pp.911-940.
4. Xia T., SantaLucia J., Burkard M.E., Kierzek R., Schroeder, S.J., Jiao X., Cox C., Turner D.H., 1998, Thermodynamic Parameters for an expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs, *Biochemistry*, 37, pp. 14719-14735.
5. Wu M., Tinoco I., Jr., 1998, RNA folding causes secondary structure rearrangement, *Proc. Natl. Acad. Sci. USA*, 95, pp. 11555-11560.
6. Kugushev, E. I., Pirogova, E. E., Starostin, E. L., 1997, A mathematical model of formation of three-dimensional structure of RNA, Moscow, Keldysh Inst. of Appl. Math., Preprint No. 77, 1997, 24 p. [in Russian].
7. Starostin E. L., 1996, Three-dimensional shapes of looped DNA, *Meccanica (Intern. J. of Italian Association of Theoretical and Applied Mechanics)*, 31, pp. 235-271.
8. Malhotra A., Gabb H.A., Harvey S.C., 1993, Modeling large nucleic acids, *Current Opinion in Structural Biology*, 3, pp. 241-246.
9. Sprinzl M., Hartmann T., Weber J., Blank J., and Zeidler R., 1989, *Nucleic Acids Research*, 17, suppl., pp. 1-172.
10. Turner D.H., Sugimoto N., 1988, RNA structure prediction, *Ann. Rev. Biophys. Biophys. Chem.*, 17, pp. 167-192.
11. Erdman V.A., Wolters J., Huysmans E., and Wachter R., 1985, *Nucleic Acids Research*, 13, pp. 105-153.
12. Benham C.J., 1983, Geometry and mechanics of DNA superhelicity, *Biopolymers*, 22, 11, pp. 2477-2495.
13. Arnott S., Chandrasekaran R., Leslie A.G.W., 1976, Structure of the single-stranded polyribonucleotide polycytidylic acid, *Journal of Molecular Biology*, 106, pp. 735-748.
14. Kim S.H., Suddath F.L., Qugley G.J., McPherson A., Sussman J.L., Wang A.H.J., Seeman N.C., Rich A., 1974, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science*, 185, 4149, pp. 435-440.
15. Love A.E.H., 1927, *A Treatise on the Mathematical Theory of Elasticity*, Cambridge Univ. Press.
16. Lewin B., 1999, *Genes VII*, Oxford Univ. Press.