ELASTIC ROD MODEL OF RNA 3D STRUCTURE

E.E.Kozyreva
State Research Institute of Genetics and Selection of Industrial Microorganisms, Moscow
E.I.Kugushev, E.L.Starostin*
M.V.Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow

*Current address: Swiss Federal Institute of Technology, Lausanne

INTRODUCTION

The spatial shape of biological molecules is known to be one of the important determinants of their biochemical properties [6]. 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 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) modeled 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 the description of the large-scale structure of DNA [2,5,10]. The 3D structure of a RNA molecule is described as a set of linked basic structural elements with known spatial configuration.

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. 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. In particular, we can assume that, when unstressed, all the rods constitute the single-stranded helix structure of the RNA in A-form.

We consider (i) dangling ends, (ii) single-stranded fragments that only join two stems and (iii) two-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) RNA regular helices. The other single-stranded parts of the molecule (loops of various type) are treated as stressed. The forces and moments are applied only at the ends of a fragment. 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 [11]. The 3D configuration of the rod fragment is defined by the system of the equilibrium equations [7]:

$$\boldsymbol{F}'=0$$
$$\boldsymbol{M}'+\boldsymbol{t}\times\boldsymbol{F}=0$$

where $\boldsymbol{F}$ is the force and $\boldsymbol{M}$ the moment, $\boldsymbol{t}$ is the tangent to the centreline of the rod and the prime denotes a derivative with respect to the arclength parameter s, measured along the centreline. The above equations are written in the laboratory reference frame. It is assumed that the moment and the Darboux vector are related by the generalized Hooke law, i.e., the components of the moment in the material reference frame are represented by the following constitutive relation:

$$M_1=A(p-p_0),\ M_2=B(q-q_0),\ M_3=C(r-r_0)$$

Here A, B are bending stiffness coefficients of the rod and C is the torsional stiffness; $p(s)$, $q(s)$, $r(s)$ and p_0, q_0, r_0 are the projections of the curvature and the twist of the centreline of the rod on the principal axes of the strain tensor in the actual and relaxed state, respectively. In order to find the shape of the centreline in the space, it is necessary to solve an additional vector equation

$$\boldsymbol{r}'=\boldsymbol{t}$$

for $\boldsymbol{r}=\boldsymbol{r}(s)$, the radius vector of points on the centreline.

The configuration of the rod depends on six parameters: A, B, C, p_0, q_0, r_0. Given these parameters, a solution to the BVP for a fragment of the rod may be found. Thus, a spatial shape of a single-stranded basic structural element is determined by finding the solution satisfying the geometrical constraints at its ends. Self-interactions of remote parts of the loop as well as of the whole molecule (tertiary interactions) are not taken into account.

The corresponding BVP is solved numerically by the shooting method. As the boundary conditions we consider 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 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 modeling the loop. A central point is defined for each. Further, we accept a simplification that the Watson-Crick bond may be represented as a rigid constraint connecting the central points of the complementary nucleotides.

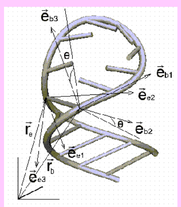

Fig. 1. Boundary value problem for the hairpin loop.

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 and 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 may be considered as a pair of interwound cross-bound rods in a relaxed state while the loop rods are in a stressed state.

At the second stage the whole structure of the molecule is assembled by means of sequential addition of the 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 to the first approximation. Therefore, all stem rods and loop rods constitute one continuous smooth rod. 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. Hence, the shape of the composite rod is completely identical to the shape that is taken on by one continuous rod of the same length which gives a solution of a multiple BVP. The latter consists of all constraints that arise due to the cross-bonds.

Another important characteristic of a loop is (the excess of) linking number. Loosely defined, it may be thought of as the number of turns to which the rod is twisted when its centreline takes on a planar shape. If we gradually rotate the right (5') end of the elastic rod by one complete turn around the tangent vector then the linking number changes to +1 or -1 depending on the direction of the rotation. Although it may be suggested that the linking number of the loop significantly affects the biological functions of the RNA and other complex biopolymers, we have found no data on the linking numbers for real loops. Hence, we choose the loop shape that has the minimal energy among all the solutions of the same BVP with different linking numbers. This particular problem deserves further investigation.

RESULTS

3D structures of RNA molecules of different types were computed by means of this technique, in particular, Yeast Phenylalanine Transfer RNA (Fig. 2a,b); its tertiary structure was determined by X-ray analysis (Fig. 2c) and described in detail in [4]. In the figures the view direction is chosen approximately perpendicular to the "plane" of the molecule.

a b c

Fig. 2. Yeast Phenylalanine Transfer RNA.

The comparison of this RNA with the result of the computation shows that even such a simple model allows one to get some qualitative resemblance of the overall conformation. Namely, the computed shape catches the following important features of the polynucleotide chain:
- 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. 3-6 are taken from [9,3].

Fig. 3. tRNA : Arginine Cenorhabdi. Elg.

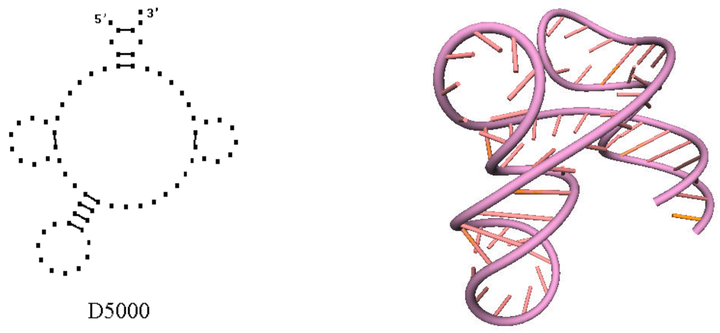

Fig. 4. tRNA : Asparagic Acid Asterina Pectini.

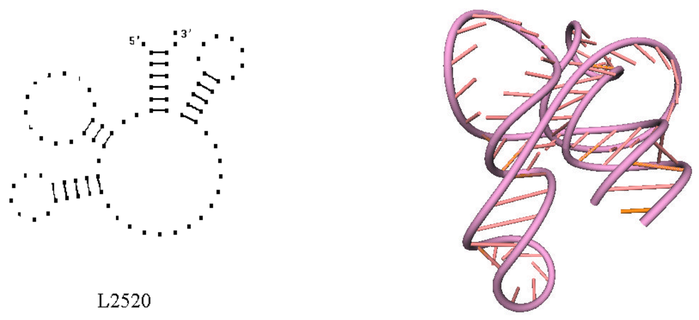

Fig. 5. tRNA : Leucine Eugelna Gracilis.

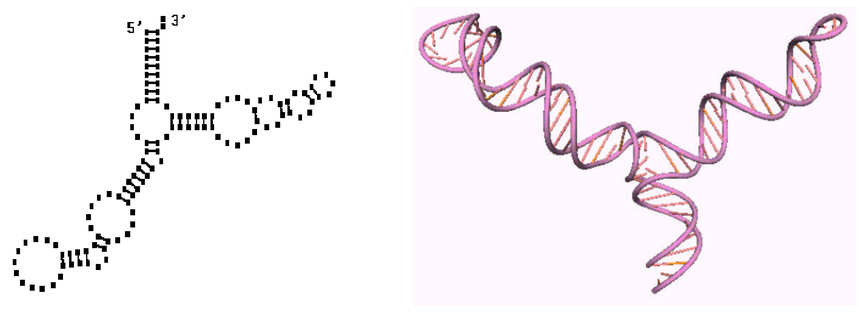

Fig. 6. 5S rRNA of human.

CONCLUDING REMARKS

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 [12]. In this respect, a means for the fast computation of a three-dimensional 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 suggested 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 [11].

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.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research under grant No. 99-01-00029.
We especially thank Prof. S.V. Mashko from the State Research Institute of Genetics and Selection of Industrial Microorganisms for his help with model improvement and 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. Benham C.J., 1983, Geometry and mechanics of DNA superhelicity, *Biopolymers*, **22**, 11, pp. 2477-2495.

3. Erdman V.A., Wolters J., Huysmans E., and Wachter R., 1985, *Nucleic Acids Research*, **13**, pp. 105-153.

4. 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.

5. Kugushev E. I., Pirogova E. E., Starostin E. L., 1997, A mathematical model of formation of three-dimensional structure of RNA, Preprint No. 77, Keldysh Inst. of Appl. Math., 24 p. [in Russian].

6. Lewin B., 1999, Genes VII, Oxford Univ. Press.

7. Love A.E.H., 1927, A Treatise on the Mathematical Theory of Elasticity, Cambridge Univ. Press.

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. Starostin E. L., 1996, Three-dimensional shapes of looped DNA, *Meccanica*, **31**, pp. 235-271.

11. Turner D.H., Sugimoto N., 1988, RNA structure prediction, *Ann. Rev. Biophys. Biophys. Chem.*, **17**, pp. 167-192.

12. Wu M., Tinoco I., Jr., 1998, RNA folding causes secondary structure rearrangement, *Proc. Natl. Acad. Sci. USA*, **95**, pp. 11555-11560.

More structures together with 3D pictures in VRML format may be found at www.geocities.com/CollegePark/Hall/3826