ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 2, pp. 236-244. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © L. B. Pereyaslavets, M. V. Baranov, E. I. Leonova, O. V. Galzitskaya, 2011, published in Biokhimiya, 2011, Vol. 76, No. 2, pp. 287-296.

Prediction of Folding Nuclei in tRNA Molecules

L. B. Pereyaslavets[1], M. V. Baranov[1,2], E. I. Leonova[1], and O. V. Galzitskaya[1]*

[1]*Institute of Protein Research, Russian Academy of Sciences, Institutskaya ul. 4, 142290 Pushchino, Moscow Region, Russia; fax: (4967) 318-435; E-mail: ogalzit@vega.protres.ru*
[2]*Biological Faculty, Lomonosov Moscow State University, 119991 Moscow, Russia*

Received June 2, 2010
Revision received September 16, 2010

Abstract—Prediction of folding nuclei in RNA molecules allows one to look in a new way at the problem of possible RNA base sequence folding and at problems associated with incorrect RNA folding, as well as at RNA structure stability. We have chosen a model and energy parameters for description of RNA structure. The algorithm for studying processes including protein folding/unfolding was successfully applied to calculations on tRNA. Four tRNA molecules were considered whose structures were obtained in the free state ($tRNA^{Phe}$, $tRNA^{Asp}$, $tRNA^{fMet}$, and $tRNA^{Lys}$). The calculated Φ-values for tRNA molecules correlate with experimental data showing that nucleotide residues in the D and T hairpin regions are involved in tRNA structure last, or more exactly, they are not included in the tRNA folding nucleus. High Φ-values in the anticodon hairpin region show that the nucleus of tRNA folding is localized just in that place.

DOI: 10.1134/S0006297911020106

Key words: dynamic programming, folding nucleus, hydrogen bond, stacking and hydrophobic interactions, coarse-grained structural model, tRNA folding

Despite simple primary structure, RNA molecules are involved in many biological processes. Spatial structure and folding of RNA molecules are now the subject of investigations of many researchers [1, 2]. Our special interest is attracted by the kinetics of this process per se and by the possibility to use nucleotide sequence for prediction of RNA secondary and tertiary structures. The latter is especially important because it is difficult to crystallize RNA molecules. Functions of RNA molecules are defined by its spatial structure, whose folding is regulated by numerous factors making RNA very similar to proteins.

Drawing the analogy between RNA and proteins, one should note that Levinthal's paradox is valid for both, according to which they find the native structure without looking over all possible conformations of polypeptide and nucleotide chains [3]. In the course of folding, the RNA strands, like a protein globule, passes through numerous intermediate states able to play the key role in the kinetics of the process.

The primary structure of biologically active RNA is formed after multiple posttranscriptional modifications and often differs from the initial transcript. These modifications include nucleotide methylation at the ribose 2′-hydroxyl group, formation of modified bases such as pseudouridyl and dihydrouridyl, and nucleotide excision or insertion [4].

The tRNA molecules do not differ much in nucleotide sequence, and their length varies from 74 to 95 nucleotides; for mitochondrial tRNA the length varies from 50 to 60 nucleotides. The same numbering of nucleotide residues is generally accepted for tRNA molecules and they are characterized by the presence of conservative (identical) sequences localized in the same positions. There are places where only purine or only pyrimidine bases, called semiconservative, are localized. Another pronounced feature of tRNA is the presence of several tens of modified (minor) bases. They are formed at definite tRNA positions in response to posttranscriptional modifications catalyzed by special enzymes. The most frequent are ribothymidine (T), pseudouridine (Ψ), 5,6-dihydrouridine (D), inosine (I), etc.

The secondary structure of all tRNA resembles a "clover leaf" and is formed by [4]:

– dihydrouridylic (D) hairpin containing dihydrouridyl residues differing by length in different tRNA;

– thymidyl-pseudouridyl (TΨ) hairpin with a stable sequence in the GTΨCGA or GTΨCAA loop;

– anticodon (AC) hairpin where all tRNA are characterized by the presence of just seven nucleotides in the loop;

* To whom correspondence should be addressed.

– acceptor (AA) stem with universal CAA sequence at its 3′ end;

– variable (V) loop strongly varying in length in different tRNAs.

All paired hairpin regions generate the A form double helix.

Tertiary structure is formed due to interaction between secondary structure elements. Little was known about this level of RNA organization until crystals were obtained for investigation by X-ray and NMR analyses. This made it possible to obtain high resolution structures of individual RNA molecules such as tRNA, the P4-P6 domain of *Tetrahymena* ribozyme, etc.

Not all nucleotides play a decisive role in RNA folding. On one side, this explains how RNA with rather low homology can form a similar tertiary structure such as tRNA, group I and II introns, and many others. On the other side, how can the removal of a single nucleotide influence the rate of RNA folding? Thus, knowledge of folding nuclei (folding nucleus is the structured part of the molecule in a transition state) makes it possible to reveal structural elements limiting the rate of RNA folding.

The tRNA molecules play an important and variable role within cells. However, the main role of tRNA is binding to its amino acid residue (with involvement of aminoacyl-tRNA synthetase) and the necessary codon recognition on mRNA. Now there are just a few experimental works on Φ-value determination for tRNA nucleotides [5]. Φ is the measure of the involvement of a residue in transition structure formation [6, 7]. If Φ = 1, this means that contacts that define the native state at the moment of the transition state have already been formed, and the given residue is incorporated in the folding nucleus. If Φ = 0, this means that these contacts are involved at the last moment of structure folding, after getting over the free-energy barrier (section "Methods of Investigation").

The authors of work [5] studied a transition state of tRNA during its unfolding. *Artificially synthesized* $tRNA^{Phe}$ was chosen for experiment, and Fersht's method for investigation of protein folding [6, 7] was used. The procedure includes insertion of point mutations into the studied molecule and detection of amino acid residues (in the case of RNA this concerns nucleotide residues), whose replacement influences stability of the transition state. Depending on the site of disulfide bond introduction (Fig. 1), it is possible both to destabilize the transition state by increasing unfolding rate and by native state stabilization. Accordingly, if a disulfide bridge joins a pair that is disrupted before or during the transition state in the course of unfolding, then unfolding rate will be decreased. A disulfide bond between a pair formed during unfolding after the transition state has no effect on unfolding rate. Therefore, several experiments on $tRNA^{Phe}$ unfolding were carried out to reveal contacts established at the last moment of folding. This was first done after introduction of a disulfide bond between nucleotides U1 and U72 in AA stem (Fig. 1, site 1). Then

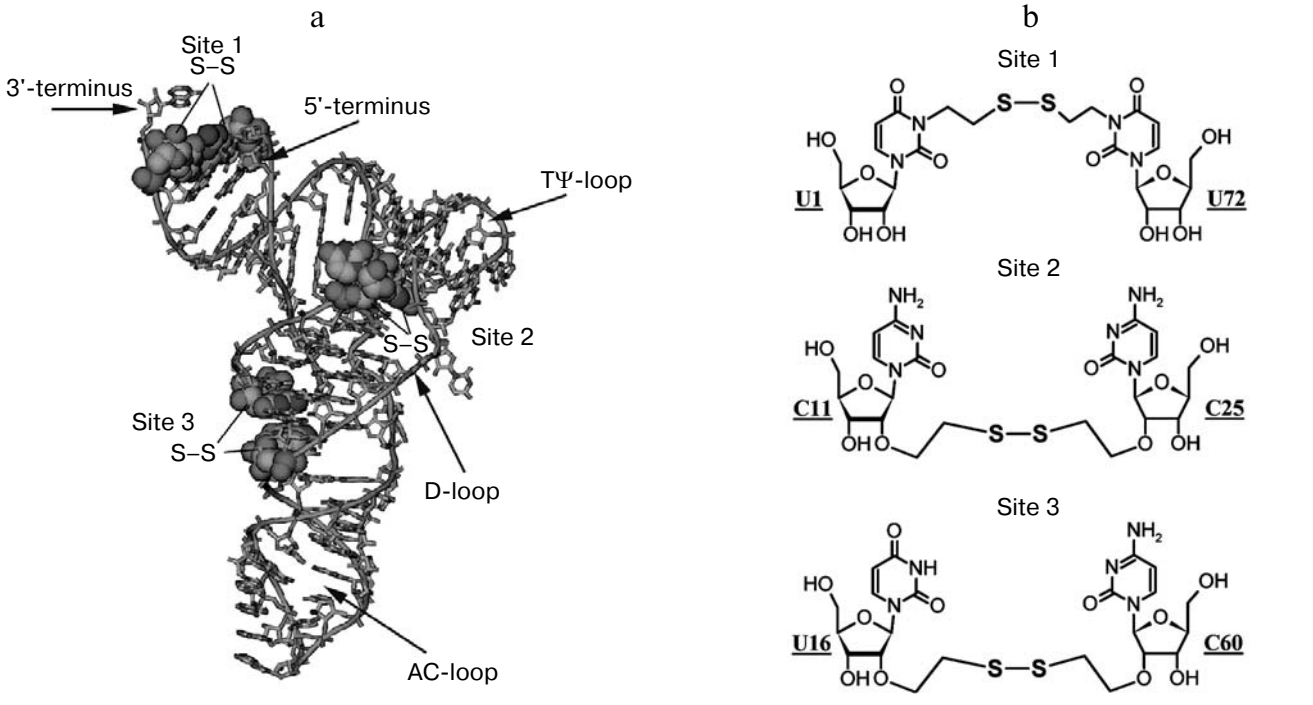

Fig. 1. a) The $tRNA^{Phe}$ molecule with indicated disulfide bonds. Site 1, disulfide bond between nucleotides U1 and U72; site 2, between C11 and C25; site 3, between U16 and C60. b) Detailed figure with introduced disulfide bonds.

a disulfide bridge was introduced between C11 and C25 nucleotides of D hairpin (Fig. 1, site 2). The last disulfide bridge joined U16 of D loop and U60 of TΨ loop (Fig. 1, site 3). The difference in free energy ($\Delta G_{F-\#}$) between the folded $tRNA^{Phe}$ molecule and the transition state was calculated for each case. Then $\Delta G_{F-\#}$ value of the molecule containing a disulfide bond was compared with $\Delta G_{F-\#}$ of the wild-type molecule (free of this bond). Unfolding was investigated at 20-35°C, which is below the tRNA melting temperature. Results of these experiments showed that $\Delta G_{F-\#}$ values do not change after disulfide bond introduction into the AA stem and D hairpin. In the case of the disulfide bond between the D and TΨ loops the difference between $\Delta G_{F-\#}$ values was about 0.8 kcal/mol. In this case the $tRNA^{Phe}$ containing the disulfide bond was unfolded in two phases, the first being temperature dependent and corresponded to melting of tertiary structure, while the second was temperature independent and indicative of slight configurational alterations following melting of the tertiary structure. Based on the obtained data, the authors concluded that contacts between D and TΨ loops are formed last during $tRNA^{Phe}$ folding. This may be due to strong repulsion between negatively charged phosphate groups. Moreover, the authors believe that a similar order of tertiary structure formation can be specific of different tRNA molecules.

It was supposed for a long time that tRNA folding (unfolding) is a hierarchical process: secondary structure is formed first, and only then contacts are formed stabilizing tertiary structure. But it has been recently shown that tRNA unfolding is not a hierarchical process [8]. It appears that consecutive unfolding of tertiary structure followed by unfolding of secondary structure is not correct for tRNA. On the whole, very often the loss of contacts associated with tertiary structure happens simultaneously with breaking base pairs of secondary structure elements [8]. This result was obtained using the specially elaborated SHAPE method [8] that provides information concerning a particular nucleotide, or more exactly, concerning its local environment of different components of the tRNA molecule. This allows one to trace the process of RNA molecule unfolding.

The fact that tRNA folding/unfolding is not hierarchical makes it possible to use a method and algorithm elaborated for protein structures for investigation of tRNA folding. Know-how of theoretical prediction of tRNA nucleotides important for formation of folding nucleus would define the most probable kinetic pathway of folding. This in turn makes it possible to schedule RNA-engineering works on experimental detection of nucleus of RNA structure folding. In this work we consider a model of RNA structure and algorithm for prediction of RNA folding nuclei. Analysis of Φ-values for four tRNA molecules whose structures were obtained in the free state ($tRNA^{Phe}$, $tRNA^{Asp}$, $tRNA^{fMet}$, and $tRNA^{Lys}$) makes it reasonable to suppose that the anticodon hairpin is incorporated in the folding nucleus, while nucleotide residues in the region of D- and T-hairpin loops are the last to be involved in tRNA structure.

METHODS OF INVESTIGATION

Assignment of coarse-grained structural model. Methods of structure representation using coarse-grained model allow one to simulate multiple molecular processes including folding or unfolding of different polymer molecules and describe the free-energy landscape of molecule folding kinetics. Go potentials [9] are usually used in these cases. This approach is now more often used for investigation of protein folding.

We have adapted the coarse-grained RNA model elaborated in Dokholyan's group [10] for description of RNA structure and energy parameters. To simplify calculations, those authors considered the full-scale atomic RNA model in which three points correspond to each nucleotide: P, phosphate group mass center; S, ribose mass center; B, mass center of the base of the six-membered ring for both purines and pyrimidines (Fig. 2).

Assignment of energy parameters for hydrogen bond, stacking, and hydrophobic interactions. The full energy of interaction in an RNA structure includes the sum of energies:

$$E_{sum} = E_{HB} + E_{Stack} + E_{Hydrophobic}, \quad (1)$$

where E_{HB} is energy of hydrogen bonds between nucleotides, E_{Stack} is energy of the nucleotide stacking interaction, and $E_{Hydrophobic}$ is the energy of hydrophobic interactions. A hydrogen bond is formed when a hydro-

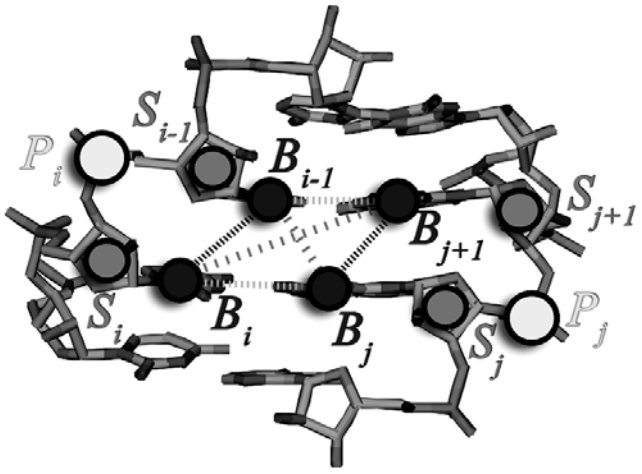

Fig. 2. Coarse-grained RNA structural model. Globules designate: B, nucleic base; S, ribose; P, phosphate. Distances vary depending on the type of nucleic base in the nucleotide. Hydrogen bonds upon interactions of the bases. Pairing contacts are shown between bases B_{i-1}:B_{j+1} and B_i:B_j.

Table 1. Distances between bases (C, G, U, and A) and sugars (S)

Nucleotide pair and its components	d_{min}	d_0	d_1	d_{max}
C_i G_j	5.20 Å	5.46 Å	5.62 Å	5.74 Å
S_i G_j	7.70 Å	8.08 Å	8.63 Å	9.00 Å
C_i S_j	9.74 Å	9.74 Å	10.53 Å	10.82 Å
A_i U_j	5.00 Å	5.25 Å	5.68 Å	5.84 Å
S_i U_j	9.76 Å	9.94 Å	10.50 Å	10.76 Å
A_i S_j	7.72 Å	7.92 Å	8.82 Å	9.00 Å
U_i G_j	5.10 Å	5.65 Å	6.10 Å	6.25 Å
S_i G_j	7.00 Å	7.44 Å	8.24 Å	8.70 Å
U_i S_j	9.50 Å	10.25 Å	10.80 Å	11.35 Å

gen atom (donor) is bound to an electronegative nitrogen or oxygen atom (acceptor). The bond strength depends on the atom size and charge. According to the model described in [10], only (A–U), (G–C), and (U–G) pairs are involved in hydrogen bonding; the basic energies of their hydrogen bonds are $\varepsilon_{HB} = -0.5$, $\varepsilon_{HB} = -1.2$, and $\varepsilon_{HB} = -0.5$ kcal/mol, respectively. If the distance between bases B_i and B_j is within the limits of d_{min}-d_{max}, then the hydrogen bond energy is calculated. The hydrogen bond energy depends on three distances (Table 1) between bases and sugars B_iB_j, S_iB_j, and B_iS_j (Fig. 2). If all these distances are within the limits of $d_1 < R < d_{max}$, then coefficient 3 is given to the hydrogen bond energy ($3\varepsilon_{HB}$). In the case of further reduction of the distance (i.e. in the case of the interval $d_0 < R < d_1$), $0.5\varepsilon_{HB}$ (kcal/mol) is added to each pair (B_iB_j, or S_iB_j, or B_iS_j). Thus, when all three distances fall within $d_{min} < R < d_0$, then $E_{HB} =$ 0 kcal/mol. When a branched hydrogen bond is formed, the energy value is divided by two.

Energy of stacking interaction decreases in the order: purine–purine → pyrimidine–purine → pyrimidine–pyrimidine [11]. Stacking formation involves quadrupole–quadrupole interactions and rather strong dispersion interaction between nucleotide bases. A certain role is also played by hydrophobic forces promoting interaction of bases, thus decreasing their interaction with solvent [11].

In our model the energy of stacking and hydrophobic interactions is considered as follows:

– if two bases are at distance $r < 4.65$ Å for purines, $r < 4.60$ Å for pyrimidines, and $r < 3.8$ Å for purine–pyrimidines as well as for all modified bases, then $E_{Stack} = -0.6$ kcal/mol;

– if base pairs are closer to each other than 6.5 Å but no stacking is formed, then the energy of hydrophobic interactions $E_{Hydrophobic} = -0.4$ kcal/mol is attributed to them. We have calculated average free energy values for these interactions from work [10].

It should be noted that for noncanonical pairs, energy of stacking and hydrophobic interactions was also considered.

Estimation of free energy. The process of consecutive folding/unfolding of native structure of nucleotide chain consisting of U nucleotide units is shown in Fig. 3. This chain has the fully folded native state S_0, fully unfolded state S_U, and multiple intermediate partially unfolded structures S_v including v disordered units and the native-like globular part of $U - v$ units ($v = 0$ for native state S_O, $v = U$ for fully unfolded state S_U, $v = 1, ..., U - 1$ for partially unfolded structures).

Unfolding of nucleotides from the chain is deliberate exclusion of any nucleotide residue from all interactions within the molecule. The removed nucleotide residue acquires the unfolded state entropy except the portion spent for closing disordered regions of the nucleotide chain. The other nucleotide residues retain their native positions. Unfolded regions do not fold into other structures different from the native one. This results in a network of folding/unfolding pathways having one fully folded structure, one fully unfolded, and multiple intermediate structures.

All free energy calculations given in this work concern the point of thermodynamic equilibrium between native structure S_0 and random coil S_U.

The free energy of an intermediate state of an RNA molecule is calculated using the equation:

$$F(S_v) = E_{sum}(S_v) - RT\,[\sigma N_{free.nucl} + S_{loop}]. \qquad (2)$$

Total energy $E_{sum}(S_v)$ is taken from all nucleotide residues of RNA structure S_v and is calculated as the sum

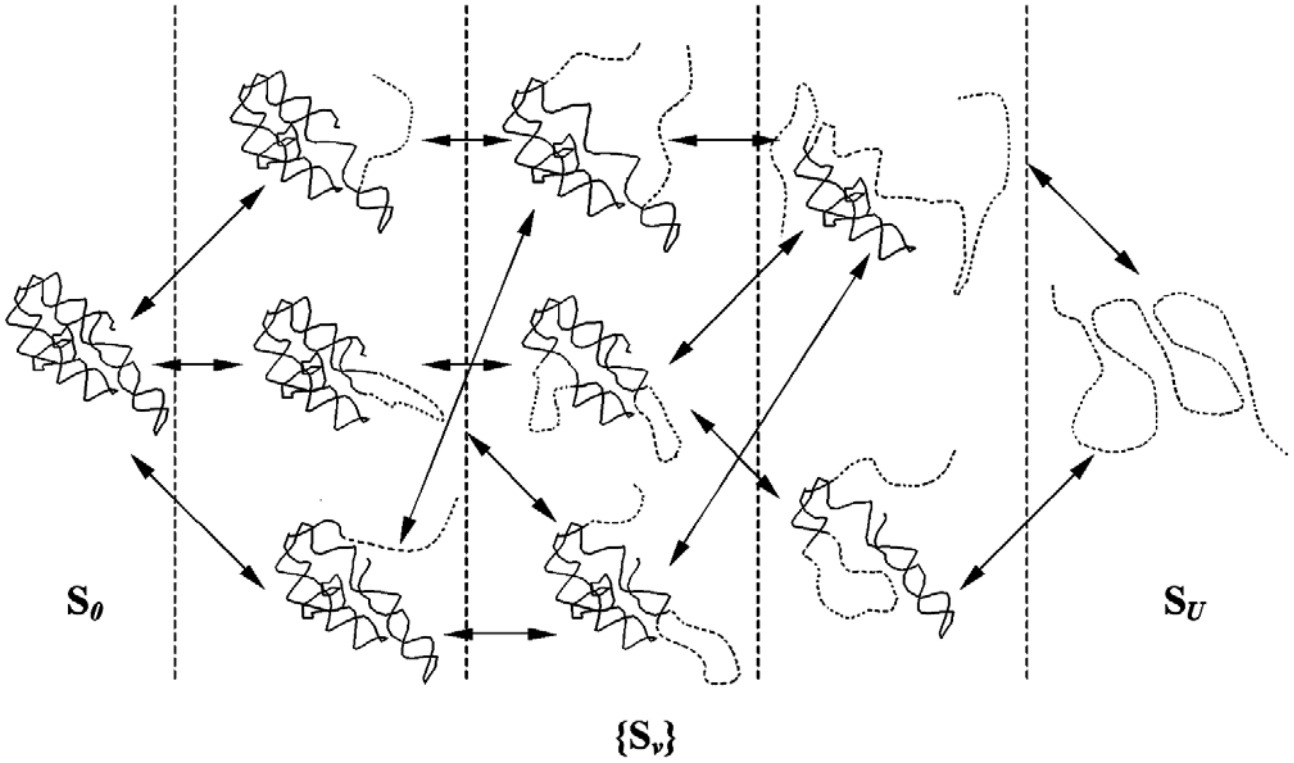

Fig. 3. Scheme of folding and unfolding pathways in native spatial structure S_0. S_U, fully unfolded state in which all U of nucleotide chain units are unfolded (this figure shows structure of domain P4-P6 from the *Tetrahymena thermophila* ribozyme first group intron). In each partially unfolded structure (type S_v), v units are unfolded (dotted line), while the other $U - v$ units retain native position and conformation (continuous line). Vertical dotted lines separate microstates with different number v of unfolded units in the chain. The central structure in the bottom row represents the microstate with v unfolded units forming one closed disordered loop and one unfolded tail; the central structure in central row is the microstate in which v unfolded units form two closed disordered loops. The pathway networks used in calculations are much more extensive than in this scheme: they include millions of partially unfolded microstates.

of energies of hydrogen bond, stacking, and hydrophobic interactions of each of nucleotides described by the coarse-grained model:

$$E_{\text{sum}}(S_v) = E_{\text{HB}} + E_{\text{Stack}} + E_{\text{Hydrophobic}}, \quad (3)$$

where E_{HB} is energy of internucleotide hydrogen bonds, E_{Stack} is energy of nucleotide stacking interactions, and $E_{\text{Hydrophobic}}$ is energy of hydrophobic interactions between nucleotides. The main designations are as follows: T, temperature in Kelvin (350 K); R, universal gas constant; σ, difference in entropy upon transition of one nucleotide residue from unfolded to structured part of the molecule in R units; $N_{\text{free.nucl}}$, number of nucleotide residues in unfolded part of the molecule; S_{loop}, Flory entropy (cost for locking loops leaving the globule between residues k and l). The loop entropy is calculated using the formula:

$$S_{\text{loop}} = -\frac{5}{2} R\ln|k - l|. \quad (4)$$

The coefficient 5 occurs due to restrictions upon integration because we restrict the loop pathway only to the space $z > 0$ [12]. We have shown in the course of protein structure modeling that the term responsible for persistent length does not make a large contribution upon calculation of the loop entropy (persistent length 20 Å) [13]. Persistent length for RNA molecules is ~10-20 Å [14], and therefore we ignore this term in this work.

Special attention should be given to calculation of σ, the entropy difference between random coil and native states of a nucleotide residue that can be calculated if the RNA structure is at the point of thermodynamic equilibrium between native and random coil phases $F(S_0) = F(S_U)$, i.e. $E_{\text{sum}}(S_0)$ and σ obey the ratio $E_{\text{sum}}(S_0) = -RTN_{\text{all.nucl}}\sigma$, where $N_{\text{all.nucl}}$ is the number of nucleotide residues in the native RNA structure.

Complete analysis of pathways passing through these "semi-unfolded" structures is carried out using the dynamic programming technique [13].

Calculation of folding nuclei. For folding nucleus calculations we used the Φ analysis method developed by Fersht for small single-domain proteins [6, 7]. The

essence of this method is insertion of point mutations and revealing amino acid residues whose replacement alters stability of the protein transition state. The $\Delta\Delta F_{N-U}$ value is the difference of free energies between the folded and unfolded states of the wild type protein and mutant, $\Delta\Delta F_{\#-U}$ is the difference between free energies of transition and denatured states. The value of ratio $\Phi = \Delta\Delta F_{\#-U}/\Delta\Delta F_{N-U}$ is the measure of the involvement of the amino acid residue in the transition state structure formation. If $\Phi = 1$, then contacts that define native state at the moment of the transition state have been already formed; this means that this residue is incorporated into the folding nucleus. If $\Phi = 0$, then these contacts evolve at the last moment of protein folding, after overcoming the free energy barrier. It is very difficult to interpret intermediate Φ-values because they depend on many factors. Such values may show both that these contacts at the moment of transition state were formed partially, and that weak interactions between pairs could be formed or not at the moment of the transition state.

Φ-values for concrete nucleotide residue (Φ_n) are calculated using the formula:

$$\Phi_n = \frac{\sum_{S^\# \in TS} \Delta_n E(S^\#) \cdot P(S^\#)}{\Delta_n E_N}, \quad (5)$$

where summation up is carried out using the ensemble of transition states constructed by the dynamic programming technique upon construction of a complete folding/unfolding network, $\Delta_n E(S^\#)$ is the change in energy of interactions upon removal of the assigned nucleotide (n) in transition state $S^\#$, the words "nucleotide removal" mean exclusion of the latter from all interactions (this is similar to a particular amino acid residue replacement by glycine in proteins [13]); $\Delta_n E_N$ is the change in interaction energy in the native state in response to the removal of nucleotide n. It is supposed that in the unfolded state the nucleotides form no contacts, i.e. they are not involved in any interaction.

To average the values in a set of transition states ($S^\#$), Boltzmann weights are used:

$$P(S^\#) = \frac{\exp(-F^\#(S^\#)/RT)}{\sum_{S^\# \in TS} \exp(-F^\#(S^\#)/RT)}, \quad (6)$$

where ($S^\#$) is the transition state from a set of all structures in this state.

Modeling conditions: the point of thermodynamic equilibrium. A completely linear RNA conformation is initially generated in the model [10] for prediction of RNA tertiary structure. However, we work with a ready RNA structure, and its atom coordinates are represented as a PDB file [15]. First, we calculate energy of hydrogen bond, stacking, and hydrophobic interactions between "grains" describing the RNA structure. Then we simulate RNA folding and unfolding under conditions of thermodynamic equilibrium when free energies between native and denatured RNA structure are identical, while the paths of direct and reverse reaction coincide according to the principle of detailed equilibrium [16]. Calculation algorithms have already been elaborated for prediction of protein unfolding pathways [13].

When we are modeling RNA structure folding and unfolding at the point of thermodynamic equilibrium by the dynamic programming technique, we do not consider incorrectly folded structure, which may be unusual blind allies or "traps" [13] and influence RNA folding or unfolding.

Network of folding/unfolding pathways. As mentioned above, spatial structure of RNA in native state borrowed from the database of PDB spatial structures is taken as the basis [15]. The RNA folding/unfolding process is modeled as reversible unfolding of its native structure by the dynamic programming technique. We consider the network of unfolding pathways in which each pathway is a simplified virtual consecutive RNA unfolding (Fig. 3), i.e. the artificial exclusion of one or another nucleotide residue from all interactions within the molecule. The removed nucleotide residue acquires the unfolded state entropy with the exception of the portion spent for closing disordered chains. The other nucleotide residues retain their native positions. Unfolded regions do not fold to another, different from native, structure.

To simplify calculations, we restrict the number of closed disordered loops protruding from the structure (no more than two loops), and in calculations of tRNA molecules we use "units" consisting of two but not of one nucleotide residue.

RESULTS AND DISCUSSION

Search for folding nuclei. Since experimental data on folding nuclei were obtained for $tRNA^{Phe}$ [5], analysis and all calculations of folding nuclei in this work have been done for $tRNA^{Phe}$, $tRNA^{Asp}$, $tRNA^{fMet}$, and $tRNA^{Lys}$ molecules whose structures in the free state (i.e. not bound in any complex) have been obtained by X-ray analysis at different resolution (Table 2).

Most hydrogen bonds are formed between standard base pairs. In our model only hydrogen bonds for three base pairs are considered (A–U, G–C, and G–U), while hydrogen bonds for all the other base pairs are not considered. For the example of yeast $tRNA^{Phe}$, these are hydrogen bonds between D and TΨ loops: G18 and Ψ55 pair, G57 with G18 and G19, and simultaneously with Ψ55 [4].

The Φ-value profile for unmodified $tRNA^{Phe}$ of *E. coli* is shown in Fig. 4. Regions with low Φ-values corresponding to loop regions of D and T hairpins as well as regions with high Φ-values corresponding to anticodon

Table 2. Calculated energy characteristics of tRNA molecules presented in this work

PDB code (resolution)	Name and origin	Energy components (kcal/mol)				Number of interactions		
		complete energy of molecule	hydrogen bonds	stacking interactions	hydrophobic interactions	number of hydrogen bonds	number of stacking interactions	number of hydrophobic interactions
1EHZ (1.93 Å)	yeast $tRNA^{Phe}$	−127.0	−31.4	−58.8	−36.8	22	98	92
1FIR (3.3 Å)	bovine $tRNA^{Lys}$	−102.83	−5.43	−52.2	−45.2	**9**	87	113
3CW5 (3.1 Å)	*E. coli* $tRNA^{fMet}$	−116.35	−17.75	−64.2	−34.4	19	107	86
3L0U (3 Å)	*E. coli* $tRNA^{Phe}$ (unmodified)	−116.1	−29.9	−58.2	−28.0	22	97	70
3TRA (3 Å)	yeast $tRNA^{Asp}$	−115.43	−25.43	−54.0	−36.0	23	90	90

hairpin are clearly seen. These data correlate with experimental results [5] showing that in the case of $tRNA^{Phe}$ destruction the contacts joining loops of D and T hairpins are broken first, and then disruption of base pairs of D hairpin secondary structure occurs.

For each tRNA type, structures with the best resolution were chosen for analysis of folding features of various tRNA. After calculation of Φ-values for four tRNA crystals, we obtained values for each component of interaction energy, number of hydrogen bonds, number of pairs of stacking and hydrophobic interactions (Table 2), and Φ-values for each nucleotide. It is necessary to note that for $tRNA^{Lys}$ with resolution 3.3 Å we obtained the fewest hydrogen bonds, which is indicative of the poor quality of the structure (Table 2).

Figure 5 shows the Φ-value profiles for four structures: $tRNA^{Phe}$ (PDB file: 1EHZ), $tRNA^{Lys}$ (1FIR), $tRNA^{fMet}$ (3CW5), and $tRNA^{Asp}$ (3TRA). These structures were crystallized in the free state, i.e. not bound in any complex. It is seen on the profile that nucleotide residues corresponding to D and T loops have the lowest Φ-values compared to different regions of the tRNA molecule, whereas anticodon helix has the highest Φ-values, which correlates with experimental data [5]. The Φ-value profile for $tRNA^{Lys}$ is slightly different, but this is due to poor resolution for this structure (3.3 Å).

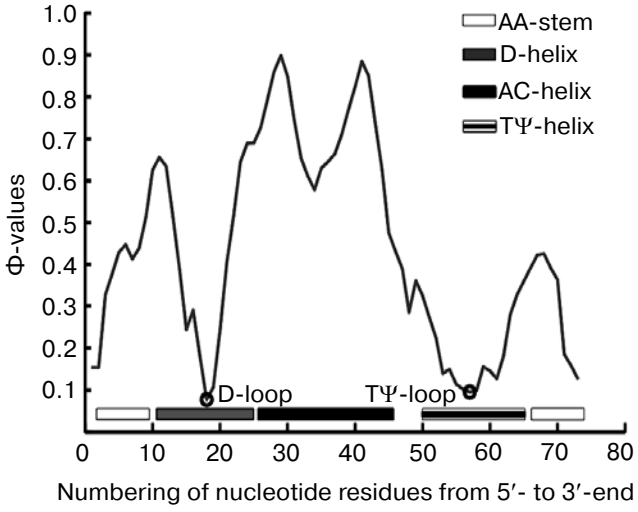

Fig. 4. Graph of *E. coli* $tRNA^{Phe}$ (PDB: 3L0U, crystalline structure of unmodified $tRNA^{Phe}$ with resolution 3.0 Å). Open circles designate nucleotides under numbers 29 and 59 corresponding to D and TΨ loops. Regions of secondary structure are marked by bars at the bottom of the figure.

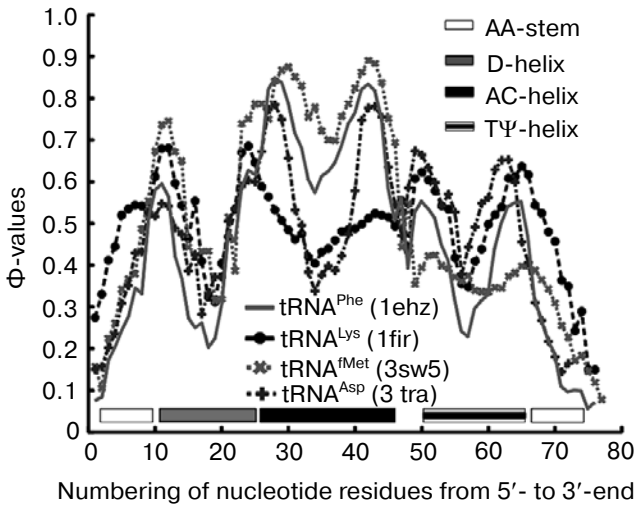

Fig. 5. Φ-value profiles for $tRNA^{Phe}$, $tRNA^{Lys}$, $tRNA^{Asp}$, and $tRNA^{fMet}$ structures: 1EHZ (with resolution 1.93 Å), 1FIR (3.3 Å), 3TRA (3 Å), and 3CW5 (3.1 Å) are corresponding PDB codes of spatial tRNA structures. Secondary structure regions are marked by bars at the bottom of the figure.

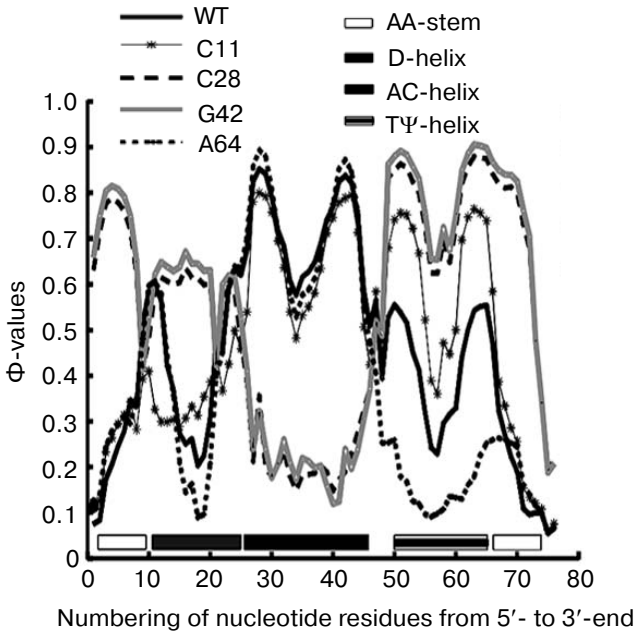

Fig. 6. Φ-value profiles for yeast $tRNA^{Phe}$ (PDB: 1EHZ). WT, wild-type line. The line with asterisks shows base removal from nucleotide No. 11 (cytosine). The fine gray line corresponds to base removal from nucleotide No. 28 (cytosine). The bold gray line corresponds to base removal from nucleotide No. 42 (guanine). The dotted line shows base removal from nucleotide No. 64 (adenine). Secondary structure regions are marked by bars at the bottom of the figure.

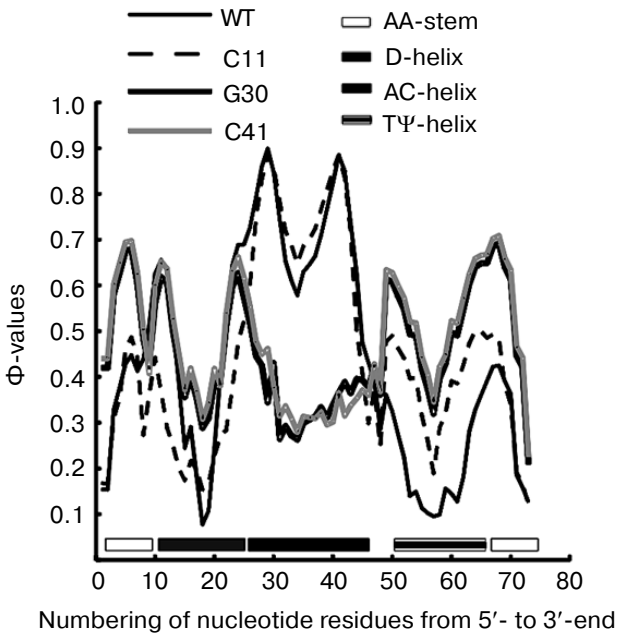

Fig. 7. Φ-value profiles for unmodified *E. coli* $tRNA^{Phe}$ (PDB: 3L0U). WT, wild-type line. The broken line shows base removal from nucleotide No. 11 (adenine). The bold black line corresponds to base removal from nucleotide No. 30 (guanine). The gray line corresponds to base removal from nucleotide No. 41 (cytosine). Secondary structure regions are marked by bars at the bottom of the figure.

Effect of particular nucleotides on localization of folding nuclei. As shown above, the folding nucleus is the structured part of the molecule in the transition state, the rate of formation of which is also able to influence the folding rate of the whole molecule. To reveal the importance of certain nucleotides in the folding nucleus, we removed some nucleotides corresponding to high Φ-values. "To remove" means to excise from the native RNA structure the base atoms of the chosen nucleotide. We suppose that the pronounced change in the profile will appear only if the removed base existed in the folding nucleus. It is well seen on the example of the $tRNA^{Phe}$ molecule (PDB file: 1EHZ) how the Φ-value profiles change after removal of G29 and C41 from the AC helix. Only minor changes were observed upon removal of base A11 from the D loop (Fig. 6). As seen in the figure, the folding nucleus is mainly localized in the anticodon hairpin region. This is also true for the *E. coli* $tRNA^{Phe}$ molecule with unmodified bases (Fig. 7). The change in the Φ-value profiles after removal of bases G30 and C41 from the AC helix is well seen in the graph (bold black and gray lines). These two cases show that a slight change in local stability of tRNA (~1 kcal/mol) is able to significantly change the pathway and, respectively, the nucleus of tRNA folding.

Theoretical work on searching for folding nuclei for tRNA molecules is of great interest. There are still very few experimental data on tRNA folding nuclei. We have successfully applied to tRNA the method of folding nuclei calculation elaborated for proteins. Theoretically calculated Φ-values for four tRNA molecules ($tRNA^{Phe}$, $tRNA^{Asp}$, $tRNA^{fMet}$, and $tRNA^{Lys}$) correspond to experimental data, indicating that nucleotide residues in the region of the D and T hairpin loops are not included in the tRNA folding nucleus. High estimated Φ-values in the region of the anticodon hairpin suggest that there is localized folding nucleus for the considered tRNA structures with resolution below 3.1 Å.

This work was supported by the Russian Foundation for Basic Research (grants No. 08-04-00561 and No. 10-04-00162), by the Russian Academy of Sciences (programs "Molecular and Cell Biology" (01200959110) and "Fundamental Sciences for Medicine"), the Foundation for Contribution to Russian Science, and the Federal Agency on Science and Innovations (02.740.11.0295).

REFERENCES

1. Woodson, S. A. (2010) *Annu. Rev. Biophys.*, **39**, 61-77.
2. Lee, M. K., Gal, M., Frydman, L., and Varani, G. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 9192-9197.

3. Levinthal, C. (1968) *J. Chem. Phys.*, **65**, 44-45.
4. Spirin, A. S. (2005) *Mol. Biol. (Moscow)*, **35**, 550-556.
5. Maglott, E. J., Goodwin, J. T., and Glick, G. D. (1999) *J. Am. Chem. Soc.*, **121**, 7461-7462.
6. Matouschek, J. T., Kellis, Jr., Serrano, L., and Fersht, A. R. (1989) *Nature*, **340**, 122-126.
7. Matouschek, J. T., Kellis, Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature*, **346**, 440-445.
8. Wilkinson, K. A., Merito, E. J., and Weeks, K. M. (2005) *J. Am. Chem. Soc.*, **127**, 4659-4667.
9. Taketomi, H., Ueda, Y., and Go, N. (1975) *Int. J. Pept. Protein Res.*, **7**, 445-459.
10. Ding, F., Sharma, S., Chalasani, P., Demidov, V. V., Broude, N. E., and Dokholyan, N. V. (2008) *RNA*, **14**, 1164-1173.
11. Saenger, W. (1987) *Principles of Nucleic Acid Structure* [Russian translation], Mir, Moscow, pp. 334-374.
12. Finkelstein, A. V., and Badretdinov, A. Ya. (1997) *Mol. Biol. (Moscow)*, **31**, 469-477.
13. Galzitskaya, O. V., and Finkelstein, A. V. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11299-11304.
14. Caliskan, G., Hyeon, C., Perez-Salas, U., Briber, R. M., Woodson, S. A., and Thirumalai, D. (2005) *Phys. Rev. Lett.*, **95**, 268-303.
15. Bernstein, F. C., Koetzte, T. F., Williams, G. J. B., Meyer, E. F., Brice, M., Rogers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *Eur. J. Biochem.*, **80**, 319-324.
16. Landsberg, P. T. (1971) *Problems in Thermodynamics and Statistical Physics*, PION, London.