Entropy **2015**, *17*, 7827-7847; doi:10.3390/e17117827

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Theoretical Search for RNA Folding Nuclei

Leonid B. Pereyaslavets and Oxana V. Galzitskaya *

Institute of Protein Research, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia; E-Mail: pereyaslavets.l@gmail.com

* Author to whom correspondence should be addressed; E-Mail: ogalzit@vega.protres.ru; Tel./Fax: +7-495-514-02-18.

Academic Editor: Kevin H. Knuth

Received: 30 April 2015 / Accepted: 17 November 2015 / Published: 23 November 2015

Abstract: The functions of RNA molecules are defined by their spatial structure, whose folding is regulated by numerous factors making RNA very similar to proteins. Prediction of RNA folding nuclei gives the possibility to take a fresh look at the problems of the multiple folding pathways of RNA molecules and RNA stability. The algorithm previously developed for prediction of protein folding nuclei has been successfully applied to ~150 various RNA structures: hairpins, tRNAs, structures with pseudoknots, and the large structured P4-P6 domain of the *Tetrahymena* group I intron RNA. The calculated Φ-values for tRNA structures agree with the experimental data obtained earlier. According to the experiment the nucleotides of the D and T hairpin loops are the last to be involved in the tRNA tertiary structure. Such agreement allowed us to do a prediction for an example of large structured RNA, the P4-P6 RNA domain. One of the advantages of our method is that it allows us to make predictions about the folding nucleus for nontrivial RNA motifs: pseudoknots and tRNA.

Keywords: RNA domain; phi value; RNA folding; mutant; pseudoknot; hairpin; stability

1. Introduction

The spatial structure and folding of RNA molecules is currently the subject of many investigations [1,2]. In the folding process, the RNA strand, like a protein globule, passes through numerous intermediate states able to play a key role in the kinetics of the process. The question of how a biopolymer chooses

its native folded form from among an astronomical number of alternative folds is acute for both RNA and protein molecules. Computer experiments with RNA-like and protein-like model chains have shown that all of them can reach their lowest-energy fold without an exhaustive search over all the possible folds. It has been shown that the possibility of fast folding of the native folds for protein-like and RNA-like heteropolymers depends on the sequence and external conditions such as temperature and solvent quality [3–6]. In trying to understand how biopolymers solve the problem of fast folding, it is useful to know the formation of what structural elements limits the RNA folding rate. It is obvious that not all nucleotides play a decisive role in RNA folding. On the one hand, this explains how RNA with rather low identity can form a similar tertiary structure such as tRNA, group I and II introns, and many others. On the other hand, how can the change of a single nucleotide influence the rate of RNA folding? Thus, knowledge about the folding nucleus (a folding nucleus is the structured part of the molecule in a transient state) makes it possible to reveal structural elements the folding of which limits the rate of RNA folding. The know-how of theoretical prediction of RNA nucleotides important for the formation of a folding nucleus would define the most probable kinetic pathway of folding. This in turn makes it possible to implement RNA-engineering for experimental detection of the RNA folding nucleus (*i.e.*, the structured part of the transition state).

The folding nucleus plays a key role in protein and RNA folding: its instability determines the folding and unfolding rate-limiting steps. It should be stressed that the folding nucleus corresponds to the free energy maximum. Therefore, there is only one, very difficult experimental method to identify the folding nuclei in proteins: to find residues whose mutations affect the folding rate by changing the transition state stability as strongly as that of the native protein [7,8]. This method is called "Φ-analysis" (see Figure 1).

Φ-values have been obtained for many proteins to find key residues in protein folding by experimental investigations [9–12]. Several theoretical algorithms have been elaborated for prediction of Φ-values in protein structures [6,13,14]. At present time there are just a few experimental works on the Φ-value determination for RNA nucleotides [15,16]. A small number of the Φ-values is available for the creation of benchmarks, but the area of RNA structure prediction is younger than that of proteins. The transition state for nucleic acid hybridization using Φ-value analysis has been tested in [17]. The authors demonstrated that the formation of a nucleation complex is a rate-limiting step, which provides an insight into effective siRNA design.

The tertiary unfolding transition state of unmodified yeast $tRNA^{Phe}$ was studied in [15]. The authors concluded that the D/T-loop junction is formed last upon the $tRNA^{Phe}$ folding. This may be due to strong repulsion between negatively charged phosphate groups. Since the tRNA tertiary structure is conserved, the authors suggest that it is possible that early unfolding of the D/T-loop junction is a common feature among tRNAs [15].

Moreover, some of the first quantitative values for an activation barrier and location of the transition state for tertiary folding of the P4-P6 RNA domain were reported in [16]. The low values of Φ indicate an early transition state for the rate-determining step of the Mg^{2+}-induced P4-P6 tertiary folding [16]. An early transition state for the P4-P6 folding is consistent with the Hammond postulate [18,19]. The low value $\Phi \sim 0$, taken alone, implies only that most of the native-state tertiary contacts are not yet formed at the rate determining folding transition state.

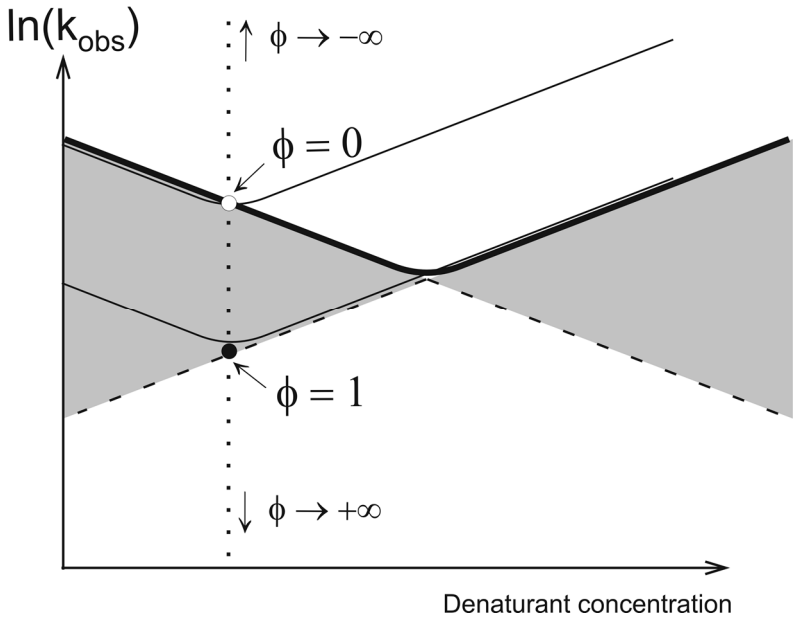


Figure 1. Scheme of the experimental identification of involvement of a residue in the folding nucleus using the Φ-analysis of site-directed mutations. The wild type chevron plot is drawn in bold. The dashed lines denote extrapolations of processes to conditions where they are non-observable in experiment (folding at high denaturant concentration and unfolding at low denaturant concentration). Closed circle: The mid-transition point for chevron plot of a mutant protein if the mutated residue has its native conformation and environment (*i.e.*, its native interactions) already in the folding nucleus; in this case $\Phi = 1$. Open circle: The mid-transition point for chevron plot of a mutant protein if the mutated residue remains denatured in the transition state; in this case $\Phi = 0$. If the mid-transition point moves from the open circle to the closed circle, the corresponding Φ-value changes from 0 to 1. If the mid-transition point moves up from the open circle, then $\Phi \to -\infty$; if the mid-transition point moves down from the closed circle, then $\Phi \to +\infty$. The grey region corresponds to the positions of mid-transition points when $0 \leq \Phi \leq 1$.

Since similarly to protein functions, RNA functions depend on the conformation of the molecule, RNA folding processes are now successfully studied using approaches developed for protein research, such as Φ analysis [20], fluorescent resonance energy transfer (FRET) [21], and small-angle X-ray scattering (SAXS) [22]. Another technique developed specially for RNA research—Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE)—allows for the identification of mobile nucleotides in RNA molecules of any size [23].

It was supposed for a long time that RNA folding (unfolding) is a hierarchical process: the secondary structure is formed first, and only then tertiary interactions are formed stabilizing the spatial structure, but using the SHAPE approach it has been recently shown that tRNA folds in a non-hierarchical manner, with non-native conformations accumulated during the folding as observed in experiments [24] and RNA secondary and tertiary interactions are formed mutually. Such a scenario of RNA folding allows us to apply the algorithm previously developed for the prediction of protein folding nuclei to the prediction of folding nuclei for RNA structures [25,26].

Theoretical studies are commonly focused on predicting the secondary and tertiary RNA structure or on describing the RNA folding kinetics presented as a free energy landscape. The investigation of the secondary RNA structure began immediately after Watson and Crick's discovery of base pairing in 1953. The RNA secondary structure, as compared to proteins, contains fewer contacts between remote chain fragments [27]. However, it is their behavior that specifies the nature of the helix-globule transition in large RNA molecules: helices formed through the pairing of remote fragments melt simultaneously by secondary phase transition, while those formed by neighboring chain fragments melt nearly independently [28]. It would be natural to suppose that the folding of large RNA molecules implying a search through an enormous number of possible structures also employs certain facilitating geometric factors. In our previous studies [29,30] it was shown that, for an RNA secondary structure model, such a factor was the presence of high energy contacts between remote chain fragments in the native structure: sequences, which were "geometrically edited" to adjust native contact energies so that the contacts between the most remote fragments were the strongest, acquired their native state in optimized external conditions (temperature and efficient component interaction) by an order of magnitude faster than random sequences.

Zuker and Stiegler's free energy minimization algorithm [31] employs the RNA sequence as input data and searches for the structure with the minimal free energy using the dynamic programming approach. The partition function algorithm is another dynamic programming algorithm suggested by McCaskill for RNA molecules without pseudoknots [32]. The partition function algorithm is included in the Vienna software package [33]. It is worth mentioning the algorithms including pseudoknots: a dynamic programming algorithm for RNA structure prediction and new server IPknot [34,35]. The major algorithm for calculating the RNA folding pathway kinetics employed the discrete molecular dynamics approach [36]. The replica exchange algorithm was used to present the energy landscape of folding pathways. Recently a coarse-grained model for predicting RNA folding thermodynamics has been developed [37].

In this work we consider a model of RNA structure adapted from the work reported in [36] and an algorithm for predicting RNA folding nuclei [25]. The analysis of Φ-values for 103 tRNA molecules whose structures were obtained in bound and unbound states 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 the tRNA structure. The calculated Φ-values for tRNA structures agree with the earlier obtained experimental data [15]. According to the experiment, the nucleotides of the D and T hairpin loops are the last to be involved in the tRNA tertiary structure.

It should be noted that many investigations have been devoted to studying and elaborating methods for predicting the folding and secondary structure for RNA molecules [27–37]. We are the first to predict the Φ-values and location of the transition states for various RNA domains. One of the advantages of our method is that we can predict the folding nucleus for any structure even with pseudoknots if its spatial structure is in the PDB or NDB databases.

2. Theory

2.1. Assignment of the Coarse-Grained Structural Model and Energy Parameters for Base Pairing, Base-Stacking, and Hydrophobic Interactions

We have developed a model similar to the coarse-grained RNA model presented by Dokholyan's group [36] to model all RNA structure and energy parameters. To simplify calculations, the authors considered a full-scale atomic RNA model in which three beads correspond to each nucleotide. Beads P and S are positioned in the center of the mass of the corresponding phosphate group and the five-atom ring sugar, the base bead (B) is positioned in the center of the six-atom ring for both purines and pyrimidines (Figure 2).

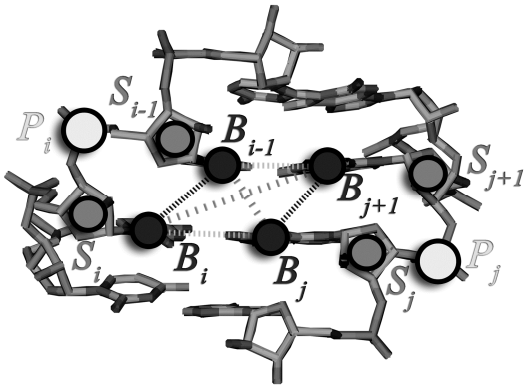


Figure 2. Coarse-grained RNA structural model. Beads in the RNA: sugar (S), phosphate (P), and base (B). Distances vary depending on the type of the nucleic base in the nucleotide. Hydrogen bonds upon interactions of the bases. Pairing contacts are shown between bases B_{i-1}:B_{j+1} and B_i:B_j.

The non-bonded interactions are crucial to model the process of RNA folding. In our adapted model [36], we have included the base pairing (only A–U, G–C, and U–G pairs are involved in hydrogen bonding), base-stacking, and hydrophobic interactions. The basic energies of the hydrogen-bonding interactions are $\varepsilon_{HB} = -0.5$ for A–U, $\varepsilon_{HB} = -1.2$ for G–C, and $\varepsilon_{HB} = -0.5$ kcal/mol for U–G, 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 (see Table 1) between bases and sugars: B_iB_j, S_iB_j, and B_iS_j (Figure 2). The distances between S_iB_j and S_jB_i, define the orientations between the two nucleotides. If the distances satisfy the predetermined range, we allow the hydrogen bond to be formed, and forbid its formation otherwise. 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). 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.

In our model, the energy of base-stacking and hydrophobic interactions is determined 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-pyrimidine as well as for all modified bases, then $E_{Stack} = -0.6$ kcal/mol; if the distance between

the base pairs is smaller than 6.5 Å but no stacking is formed, then the energy of hydrophobic interactions $E_{\text{Hydrophobic}} = -0.4$ kcal/mol is attributed to them. We calculated the average free energy values for these interactions using the data from [36]. The considered parameters for all possible neighboring base pairs are given from the experimentally tabulated energy [38]. It should be noted that for non-canonical pairs, the energy of stacking and hydrophobic interactions was also considered.

Table 1. Distances between bases (C, G, U, 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

2.2. Network of Folding/Unfolding Pathways and the Point of Thermodynamic Equilibrium

Why do we investigate the RNA unfolding rather than folding? Simulation of unfolding is simpler than that of folding because one can avoid exploring numerous high-energy dead-ends, while, according to the detailed balance principle [6], the pathways for folding and unfolding must coincide when both processes take place under the same conditions. Hence, we are interested in conditions close to that of thermodynamic equilibrium between the native and the coil states. Calculation algorithms were already developed for the prediction of protein unfolding pathways [6].

Spatial structure of RNA in the native state was picked up from the PDB or NDB databases [39]. The RNA folding/unfolding process is modeled as reversible unfolding of its native structure by the dynamic programming technique [6]. We consider the network of unfolding pathways in which each pathway is a simplified virtual consecutive RNA unfolding (Figure 3), *i.e.*, the artificial exclusion of one or another nucleotide from all interactions within the molecule. The removed nucleotide gains the unfolded state entropy with the exception of the entropy spent to close the disordered loops protruding from the remaining structure. It is assumed that the other nucleotides retain their native positions and that the unfolded regions do not fold into another, non-native structure. To use dynamic programming to search out an ensemble of transition states (or a transition state) in a large network of folding-unfolding pathways, we have to restrict this network to ~10^7 intermediates. Therefore we consider only the intermediates with no more than two closed loops in the middle of the chain plus the N- and the C-terminal disordered tails. To the same end, we use "strand links" consisting of a few nucleotides: of two for RNA with less than 80 nucleotide residues, and of four (or three) for larger RNAs.

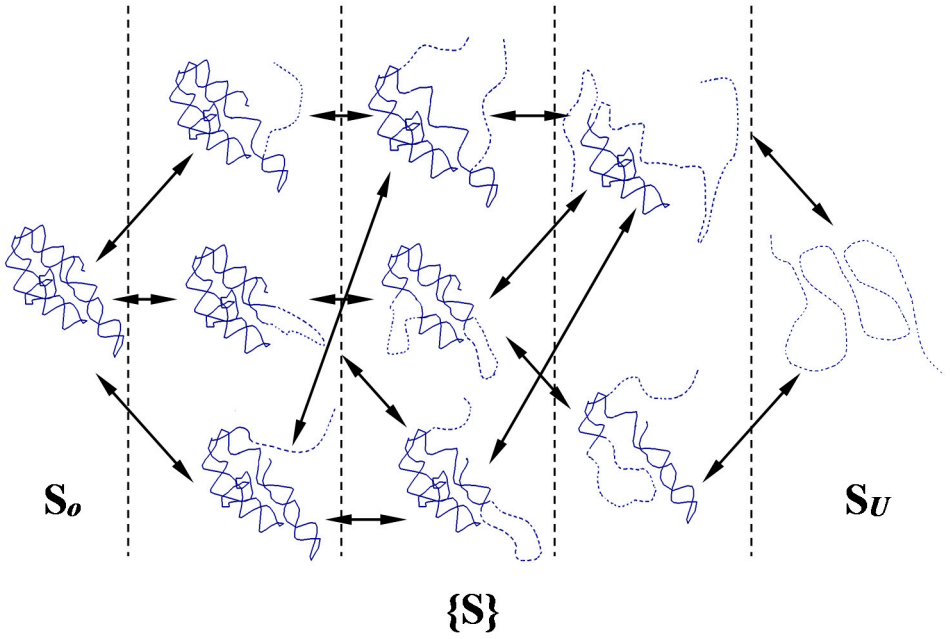


Figure 3. Scheme of folding and unfolding pathways in native spatial structure S_0. S_U, fully unfolded state U in which all nucleotide chain links are unfolded (this figure shows the structure of domain P4-P6 from the *Tetrahymena thermophila* ribozyme first group intron). In each partially unfolded structure (type S_v), v links are unfolded (dotted line), while the other $U - v$ links retain their native position and conformation (continuous line). Vertical dotted lines separate microstates with different number v of unfolded links in the chain. The central structure in the bottom row represents the microstate with v unfolded links forming one closed disordered loop and one unfolded tail; the central structure in the central row is the microstate in which v unfolded links 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.

2.3. Estimation of Free Energy and Calculation of Folding Nuclei

The process of consecutive folding/unfolding of native structure of a nucleotide chain consisting of U nucleotide links is shown in Figure 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 links and the native-like globular part of $U - v$ links ($v = 0$ for native state S_0, $v = U$ for fully unfolded state S_U, $v = 1, \ldots, U - 1$ for partially unfolded structures).

All free energy calculations given in this work relate to 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:

$$F(S_v) = E_{sum}(S_v) - RT[\sigma N_{free.nucl} + S_{loop}] \quad (1)$$

The total energy $E_{sum}(S_v)$ is calculated by taking into account all nucleotides of RNA structure S_v and is presented as the sum of energies of base pairing (energies of hydrogen bond, E_{HB}), base-stacking

(E_{Stack}), and hydrophobic interactions ($E_{\text{Hydrophobic}}$) of each of nucleotides described by the coarse-grained model:

$$E_{\text{sum}}(\text{S}_v) = E_{\text{HB}} + E_{\text{Stack}} + E_{\text{Hydrophobic}} \quad (2)$$

The main designations are as follows: T is the temperature in Kelvin (350 K); R is the universal gas constant; σ is the difference in entropy upon transition of one nucleotide residue from unfolded to the structured part of the molecule in R units; $N_{\text{free.nucl}}$ is the number of nucleotide residues in the unfolded part of the molecule; S_{loop} is the loop entropy (the cost for locking loops leaving the globule between residues k and l). As we consider the loops protruding from the globular state the Jacobson-Stockmayer model can be considered [40,41]. The loop entropy is calculated using the formula:

$$\text{S}_{\text{loop}} = -\frac{5}{2} R \ln|k - l| \quad (3)$$

The factor of 5 is due to limitations in the integration, as we restrict the movement of the loop only in space z > 0 [42]. Upon protein structure modeling we have shown that the term responsible for the persistent length does not make a large contribution to the calculations of the loop entropy (the persistent length 20 Å) [6]. The persistence length for RNA molecules is ~10–20 Å [43], and therefore here we ignore this parameter (the term itself is not so strict and can be largely treated as a constant in the first approximation [44]).

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 $F(\text{S}_0) = F(\text{S}_U)$, *i.e.*, $E_{\text{sum}}(\text{S}_0)$ and σ obey the ratio $E_{\text{sum}}(\text{S}_0) = -RTN_{\text{all.nucl}}\sigma$, where $N_{\text{all.nucl}}$ is the number of nucleotides in the native RNA structure (the chain fragment corresponding to the link includes $N_{\text{all.nucl}}/U$ nucleotides). A complete analysis of the pathways through these "semi-unfolded" structures is carried out using the dynamic programming technique [6].

The value of the ratio $\Phi = \Delta\Delta F_{\#-\text{U}}/\Delta\Delta F_{\text{N}-\text{U}}$ is the measure of the involvement of the nucleic acid in the transition state structure formation. The $\Delta\Delta F_{\text{N}-\text{U}}$ value is the difference of free energies between the folded and unfolded states of the wild type RNA and the mutant, $\Delta\Delta F_{\#-\text{U}}$ is the difference between free energies of transition and denatured states. If $\Phi = 1$, then contacts that define the 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 RNA 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 the 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 (Φ_n) are calculated using the formula:

$$\Phi_n = \frac{\sum_{S\#\in TS} \Delta_n E(S^{\#}) \cdot P(S^{\#})}{\Delta_n E_N} \quad (4)$$

where summation is made using the ensemble of transition states generated by the dynamic programming technique upon construction of a complete folding/unfolding network, $\Delta_n E(\text{S}^{\#})$ 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 [6]); $\Delta_n E_N$ is the change in the 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 transient states ($S^{\#}$), Boltzmann weights are used:

$$P(S^{\#}) = \frac{\exp(-F^{\#}(S^{\#})/RT)}{\sum_{S\#\in TS}\exp(-F^{\#}(S\#)/RT)} \quad (5)$$

where ($S^{\#}$) is the transient state from a set of all structures in this state. These Φ-values have the same sense as the Φ_f values derived from the protein/RNA engineering experiments. They are compared to see the theory and experiment correlation.

3. Results and Discussion

3.1. Prediction of Folding Nuclei for tRNAs

Do RNA and protein chains have different folding kinetics? To answer this question we investigated the transition state for the RNA folding kinetics suggesting that tertiary and secondary structures are formed simultaneously for different RNA domains/sequences [24].

Transfer RNA (tRNA) molecules play an important and variable role within cells. However, the main role of tRNA is the binding to its amino acid residue (with involvement of aminoacyl-tRNA synthetase) and the necessary codon recognition on mRNA. Although tRNA is often subject to a variety of modifications, it still serves as a valuable model experiment and even as a benchmark in studying three dimensional (3D) folding [36]. Searching for various tRNA structures we have scanned the NDB database and obtained 126 files including the structures of various tRNA molecules. Seventeen kinds of various tRNAs molecules were found among 103 tRNA structures with resolution better than 3 Å. It turns out that practically all tRNA structures (90) are crystallized in the complex with aminoacyl_tRNA synthetase or with part of ribosomal RNA (bound tRNA). Only 13 tRNA structures ($tRNA^{Phe}$, $tRNA^{Asp}$, $tRNA^{fMet}$, and $tRNA^{Lys}$) were determined in the unbound state [25]. All calculations and analysis of folding nuclei in this work were performed for these 113 tRNA structures (http://bioinfo.protres.ru/foldnucleus/Foldnucleus_tRNA.pdf).

Since experimental data on folding nuclei were obtained for $tRNA^{Phe}$ [15], we began from this structure. The Φ-value profile for unmodified $tRNA^{Phe}$ of *E. coli* is shown in Figure 4. Regions with low Φ-values corresponding to loop regions of the D and T hairpins as well as regions with high Φ-values corresponding to anticodon hairpin are clearly seen. These data agree with the experimental results [15] showing that in the case of $tRNA^{Phe}$ destruction the contacts joining the loops of the D and T hairpins are broken first, and then disruption of base pairs of the D hairpin secondary structure occurs.

Figure 5 shows the Φ-value profiles for four structures determined in the unbound states: $tRNA^{Phe}$ (NDB file: 1EHZ), $tRNA^{Lys}$ (1FIR), $tRNA^{fMet}$ (3CW5), and $tRNA^{Asp}$ (3TRA). One can see that these profiles look similar: nucleotides corresponding to the D and T loops have the lowest Φ-values compared to different regions of the tRNA molecule, whereas the anticodon helix has the highest

Φ-values, which agrees with the experimental data [15]. The Φ-value profile for $tRNA^{Lys}$ is slightly different, but this is due to poor resolution for this structure (3.3 Å). We obtained values for each component of interaction energy, the number of hydrogen bonds, the number of pairs of stacking and hydrophobic interactions for the four considered structures (Table 2).

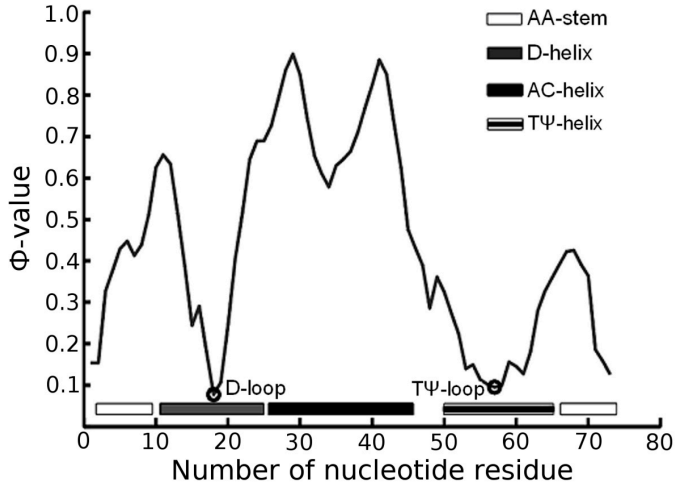


Figure 4. Graph of calculated Φ-values for *E. coli* $tRNA^{Phe}$ (PDB: 3L0U, crystalline structure of unmodified $tRNA^{Phe}$ with resolution 3.0 Å). Open circles designate nucleotides numbered 19 and 59 corresponding to D and TΨ loops. Regions of secondary structure are marked by bars at the bottom of the figure.

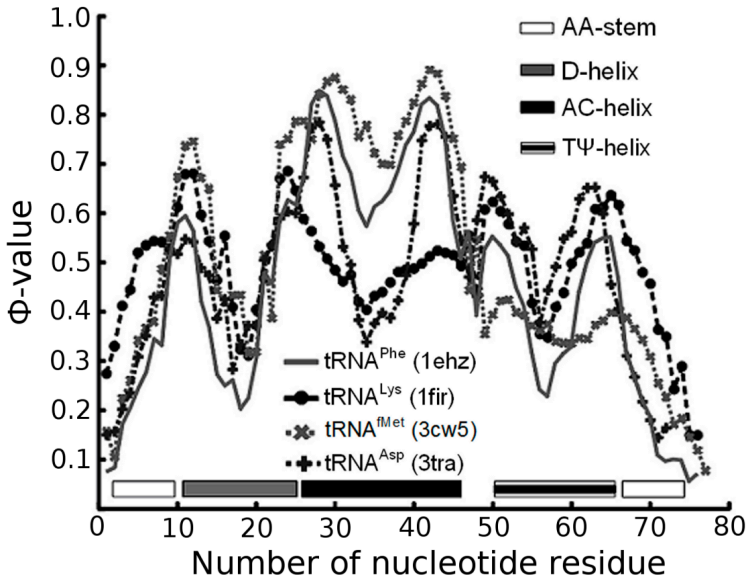


Figure 5. Predicted Φ-value profiles for $tRNA^{Phe}$, $tRNA^{Lys}$, $tRNA^{Asp}$, and $tRNA^{fMet}$ structures: 1EHZ (with resolution 1.93 Å), 1FIR (3.3 Å), 3TRA (3 Å), and 3CW5 (3.1 Å) are the corresponding PDB codes of spatial tRNA structures. Secondary structure regions are marked by bars at the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

The profiles of Φ-values for two different bound chains of $tRNA^{Glu}$ molecules are shown in Figure 6. It is seen that the predicted Φ-values in the anticodon region changed greatly. The profiles of

Φ-values for 17 types of various tRNAs molecules are presented at http://bioinfo.protres.ru/foldnucleus/Foldnucleus_tRNA.pdf. As seen, the Φ-value profiles for bound tRNA structures sometimes differ or coincide with the Φ-value profiles obtained for an unbound tRNA molecule. The analysis of literature data allows us to make some conclusions about this fact. It turns out that in most cases the crystal structures of protein-tRNA complexes reveal a range of structural rearrangements, in some cases they are dramatic, occurring upon protein binding. For example, distortion of the anticodon loop is often such that bases are presented to the aminoacyl-tRNA synthetase anticodon-binding domain for recognition, as in class I human aminoacyl-tRNA synthetase: $tRNA^{Trp}$ [PDB entry: 2DRS] [45] and class II *Escherichia coli* aminoacyl-tRNA synthetase: $tRNA^{Asp}$ complexes [PDB entry: 1C0A] [46]. Distortions of the acceptor stem are also observed, particularly for tRNAs aminoacylated by class I aminoacyl-tRNA synthetases, which approach their tRNA partners from the acceptor stem minor groove side. It should be noted here that the enzymes modifying anticodon nucleotides distort this loop to access the bases.

Table 2. Calculated energy characteristics of unbound tRNA molecules.

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 $tRNA^{Phe}$	−127.2	−31.4	−59.4	−36.4	22	99	90
1FIR (3.3 Å)	Bovine $tRNA^{Lys}$	−116.62	−21.82	−60.0	−34.8	20	100	86
3CW5 (3.1 Å)	*E. coli* $tRNA^{fMet}$	−116.75	−17.75	−67.8	−31.2	19	113	77
3L0U (3 Å)	*E. coli* $tRNA^{Phe}$ (unmodified)	−116.1	−29.9	−58.2	−28.0	22	97	69
3TRA (3 Å)	Yeast $tRNA^{Asp}$	−116.43	−25.43	−57.0	−34.0	23	95	84

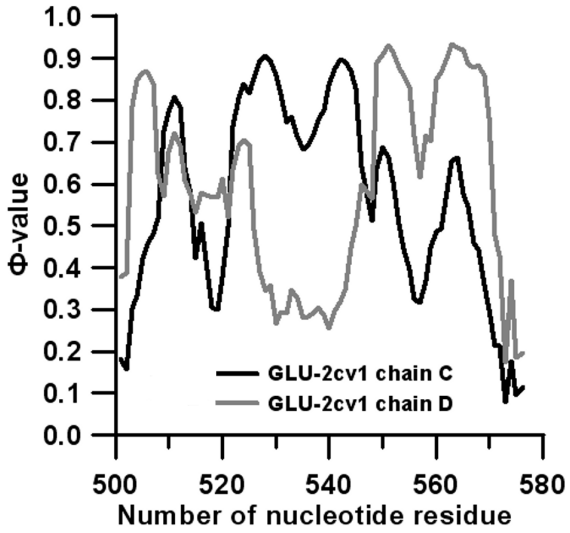


Figure 6. Calculated Φ-value profiles for two different bound chains of $tRNA^{Glu}$.

In the cases when the structure is disturbed upon complex formation, the calculation of the folding nuclei is not correct. For example, the anticodon loop of $tRNA^{Gln}$ in the complex with aminoacyl-tRNA

synthetase has an unusual structure: U35 forms base-stacking with A37, but C34 and G36 begin to interact with the side chains of the protein. Likely, such described distortions of the structures are typical for aminoacyl-tRNA synthetase belonging to the first class upon interaction with corresponding tRNA. It should be noted that in some cases (for example) the A-form of $tRNA^{Asp}$ (2TRA) can distort the RNA structure.

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 the base atoms of the chosen nucleotide from the native RNA structure. We suppose that the pronounced change in the profile will appear only if the removed base exists in the folding nucleus. It is well seen on the example of the $tRNA^{Phe}$ molecule (PDB file: 3L0U) how the Φ-value profiles change after removal of G30 and C41 from the AC helix. Only minor changes were observed upon removal of base C11 from the D loop (Figure 7). As seen in the Figure, the folding nucleus is mainly localized in the anticodon hairpin region. These two cases show that a slight change in local stability of tRNA (~1 kcal/mol) is able to significantly change the pathway and, correspondingly, the nucleus of tRNA folding. This situation is similar for tRNA molecules crystallized in the bound state.

It should be underlined here that in many protein families the transition states of homologous proteins are remarkably similar [47]. For tRNA molecules we observed a similar situation: the general structural features of the transition state are maintained.

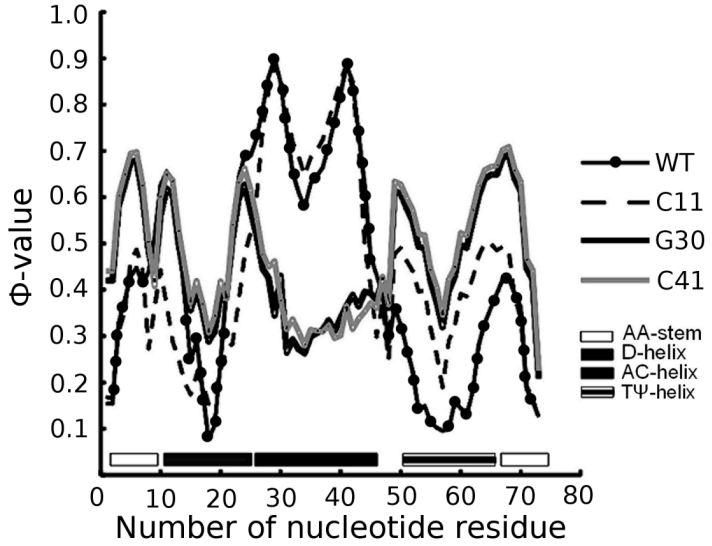


Figure 7. Calculated Φ-value profiles for unmodified *E. coli* $tRNA^{Phe}$ (PDB:3L0U). WT is a wild-type line. The broken line shows base removal from nucleotide 11 (adenine). The bold black line corresponds to base removal from nucleotide 30 (guanine). The gray line corresponds to base removal from nucleotide 41 (cytosine). Secondary structure regions are marked by bars at the bottom of the figure.

3.2. Prediction of Folding Nuclei for Domain P4-P6 from the Tetrahymena thermophila Ribozyme First Group Intron

The kinetics simulations suggest that the folding of the RNA secondary structure depends on the structural factors of the native state. It has been demonstrated that one of these factors accelerating the

folding of large RNA is the existence of strong long-range helices in the native secondary structure [29,30]. The available conformational space increases exponentially with the increasing length of the simulated RNA. Moreover, RNA is structurally very flexible and this flexibility will increase also with the increasing RNA length. Practically, all RNA based reactions are thought to take place together with proteins [48]. In this work we have made predictions of the folding nuclei for one of the largest known spatial structure of RNA, domain P4-P6 from the *Tetrahymena thermophila* ribozyme first group intron. Compared to tRNA, the 160-b P4-P6 domain is considerably larger, and the investigation of its folding represents the next stage leading to the understanding of the self-organization of large nucleic acids. This ribozyme belongs to group I introns due to the presence of several short conserved sequence elements determining its secondary structure [49]. The ribozyme comprises nine paired helical fragments P1-P9 [49].

The impact of the P1 domain on the tertiary structure formation of the full group I intron ribozyme was studied in [50]. P1 is a duplex of six base pairs. Apparently, the docking of P1 to the preformed ribozyme core is the last and therefore represents the limiting stage in the formation of a functional molecule. The experiment involved introducing eight different chemical modifications of P1 nucleotides and analyzing their effects on the correct docking of P1 to the ribozyme core by means of FRET. The Φ-values for the modified nucleotides of P1 were calculated using the docking rate and equilibrium constants k_{dock} and K_{dock}. P1 was assumed to be initially not docked to the rest of the ribozyme, *i.e.*, in a quasi-free state. In this case the Φ-values were calculated using the following equation:

$$\Phi = \frac{\Delta\Delta G(k_{dock,\mathrm{mod}ified \to un\,\mathrm{mod}ified})}{\Delta\Delta G(K_{dock,\mathrm{mod}ified \to un\,\mathrm{mod}ified})} \quad (6)$$

It was shown that the P1 duplex was the last to dock to the catalytic core (Φ = 0). The authors concluded that large RNAs in their native state may be comprised of locally unfolded fragments [50].

Using the experimental data available for domain P4-P6 it was found that it is a highly co-operative folding; it was also shown that the tertiary contacts in the particular domain P4-P6 are formed two times faster than in the same domain, but are included in the whole ribozyme. The P4-P6 domain can be studied independently from the whole molecule, since its folding is not controlled by the rest of the multi-domain structure. At the same time a three-helix junction (P5abc) within the P4-P6 domain folds at least 25 times more rapidly in isolation than when being a part of the wild-type P4-P6 RNA [51]. The fastest observed rate constants for P4-P6 folding are on the order of 10 s^{-1} [52], much slower than the formation of basic structures such as RNA hairpins, which form on the microsecond timescale [53,54].

To follow the formation of equilibrium tertiary structures, the folding kinetics of the pyrene-labeled P4-P6 domain at different Mg^{2+} concentrations was monitored using the stop-flow technique [16]. The observed folding rate constant k_{obs} was assessed by changing the fluorescence intensity at different Mg^{2+} concentrations, and the free activation energy $\Delta G^{\#}$ of Mg^{2+} induced domain folding was calculated using the formula (Eyring equation):

$$k_{obs} = (k_b\mathrm{T}/\mathrm{h})\exp(-\Delta \mathrm{G}^{\#}/k_bT) \quad (7)$$

To identify the nucleotides included in the folding nucleus and affecting the folding of P4-P6, several mutations were introduced into the domain (Figure 8). The Φ-values were determined based on

the changes in the free energies of folding and the stable state. In spite of some thermodynamic differences, the folding rates of wild-type and mutated P4-P6 were similar. At the same time, the fact that $\Phi \sim 0$ indicates that, in the transition state, these nucleotides had not yet formed their native contacts.

For instance, at the Mg^{2+} concentration of approximately 10 mM, P4-P6 folded within milliseconds with k_{obs} of 15 to 31 s^{-1}. For the given folding rate constant, $\Delta G^{\#}$ lies within the range of 8 to 16 kcal/mol. The relatively wide range obtained for free energy may be due to the uncertainty regarding the pre-exponential factor used in the formulas for k_{obs} and $\Delta G^{\#}$. In the Eyring equation used in this case, the pre-exponential factor was suggested for small molecule reactions in the gas phase. However, the P4-P6 molecule is fairly large, and the use of the k_BT/h factor may be questionable [16]. In spite of this problem, the activation energy values obtained in the study look fairly reasonable, as they agree with the limit of the known free energy barrier of tertiary folding of protein and RNA molecules (typically, a few kcal/mol).

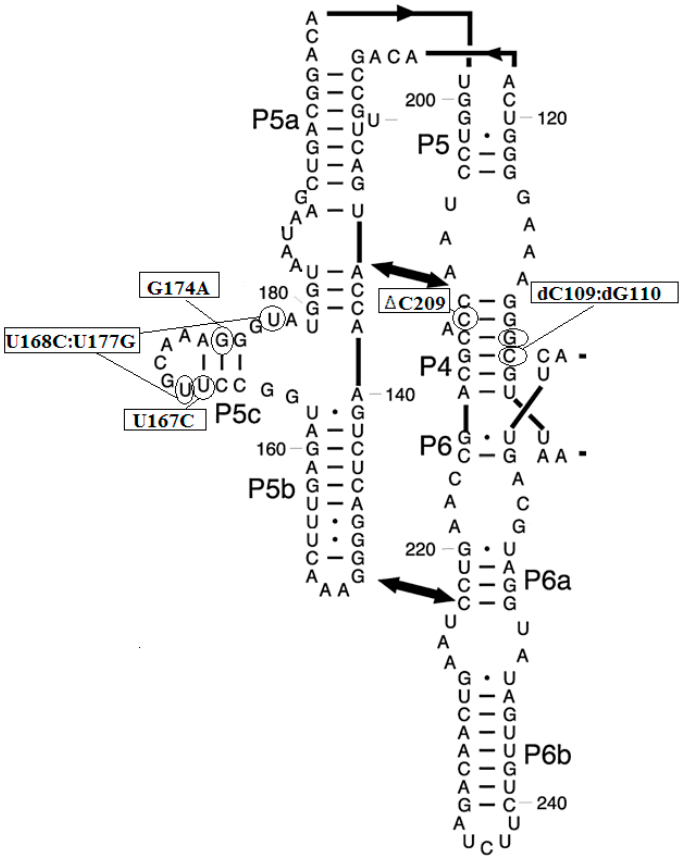


Figure 8. Secondary structure of the P4-P6 domain and the sites of mutations introduced.

As can be seen from Figure 9, our theoretically calculated Φ-values for domain P4-P6 are not in close agreement with the meager experimental data [16], but differences between our theoretical assumptions and the experimental conditions may be largely to blame for these discrepancies. As it turned out, the mutations (marked with yellow dots in Figure 9) were produced mainly by the nucleotides involved in the formation of tertiary contacts and gave the lowest Φ-values in the experiment. Moreover, in the simulations aimed at estimation of the folding nuclei in RNA molecules, it is important to take into account its state (fully unfolded or just losing tertiary contacts). The experiment was carried out determining the Φ-values only for the nucleotides involved in tertiary

contacts (the molecule retained its secondary structure); and in our theoretical study, we expect the Φ-values of the completely unfolded state. Perhaps that is why there is such a discrepancy between the theoretically calculated and experimental Φ-values in the 5 nucleotide residues. Another reason is a small thermodynamic effect of mutations.

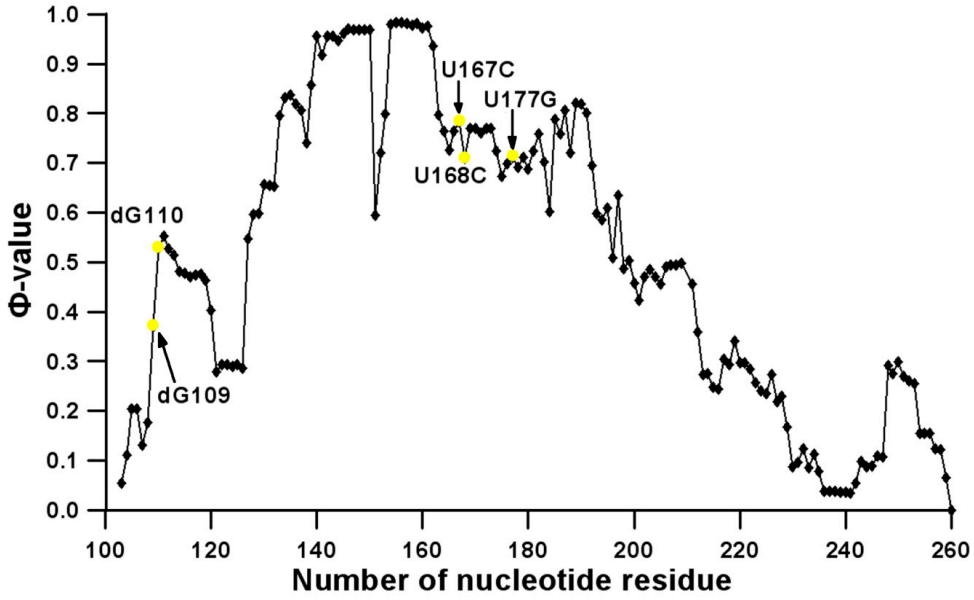


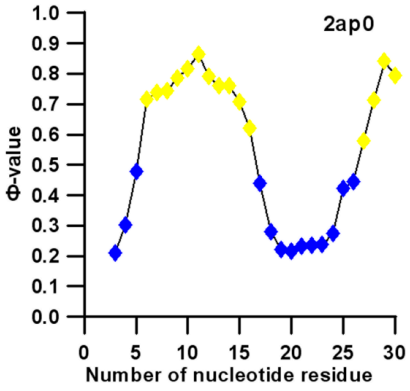
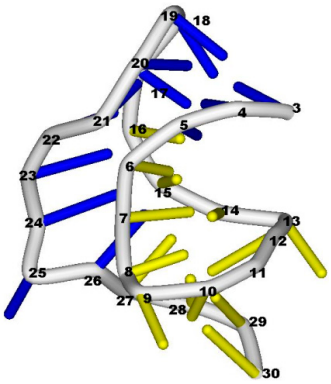
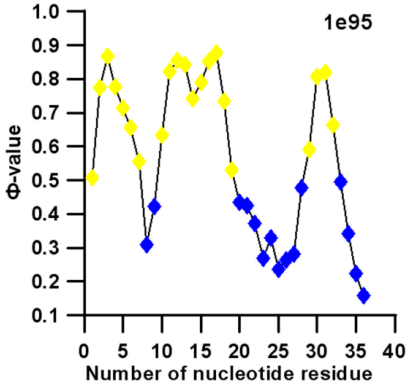
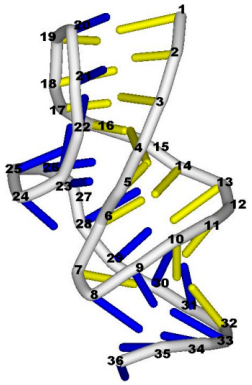
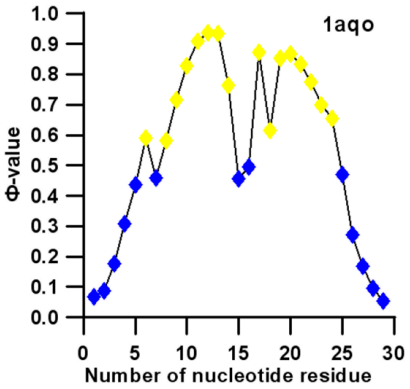
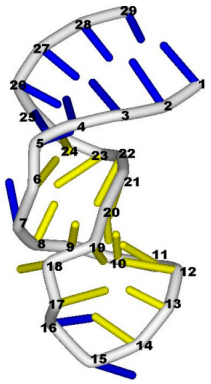
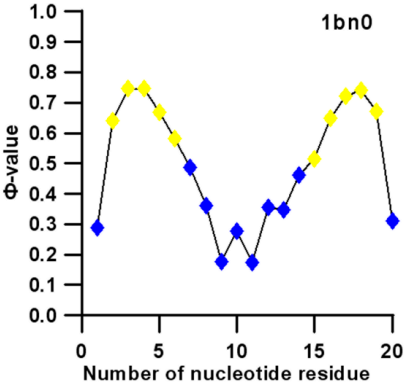
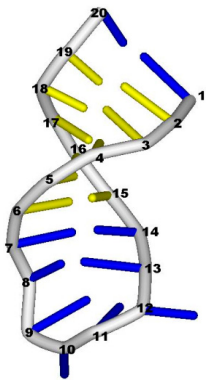
Figure 9. Theoretically predicted Φ-values for domain P4-P6 from the *Tetrahymena thermophila* ribozyme first group intron. Yellow circles point out the nucleotides for which experimental measured Φ-values are close to zero.

It was reported that Φ-values were considered reliable only for protein mutations whose effect on the stability of the native state exceeded 1.6 kcal/mol [55]. From the experimental data we have for P4-P6 that ΔC209 deletion stabilizes folding of P4-P6 [56] by 1.1 kcal/mol, and the estimated $\Delta\Delta G$ is 0.4 kcal/mol for the destabilizing U168C:U177G double mutation [57]. The dual observations of a large $\Delta G^{\#}$ value and an early transition state (low Φ) suggest that a simple two-state energy diagram for the P4-P6 tertiary folding is incomplete [16]. The conclusion of the authors is that it is necessary to calibrate the relationship between k_{obs} and $\Delta G^{\#}$ for all RNAs, including P4-P6.

3.3. Prediction of Folding Nuclei for RNA Structures with Hairpin and Pseudoknots

An RNA pseudoknot is minimally comprised of two helical structures connected by two single-stranded loops, thereby providing a simple way in which a single-strand of RNA can fold back on itself. As such, RNA pseudoknots are widely recognized to perform diverse biological functions, including the formation of protein recognition sites mediating replication and translational initiation, self-cleaving ribozyme catalysis, and inducing frameshifts in ribosomes [58]. It should be noted that many studies have been devoted to studying and elaborating the methods of predictions for folding and secondary structure for RNA molecules without a pseudoknot. One of the advantages of our method is that we can predict the folding nucleus for any structures even with pseudoknots if there is its spatial structure in the PDB or NDB databases. We study pseudoknots folding/unfolding by selecting representative pseudoknots whose structures are available in the PDB or NDB (see Table 3).

Table 3. Profiles of Φ -values for RNA pseudoknots and hairpins structures. Yellow color corresponds to nucleotides with high probability to be in the folding nucleus.

PDB Entry	Name	Φ -value Profile	3D Structure of Molecule
2ap0 (NMR)	C27A SUGARCANE YELLOW LEAF VIRUS RNA PSEUDOKNOT		
1e95 (NMR)	SOLUTION STRUCTURE OF THE PSEUDOKNOT OF SRV-1 RNA, INVOLVED IN RIBOSOMAL FRAMESHIFTING		
1aqo (NMR)	IRON RESPONSIVE ELEMENT RNA HAIRPIN		
1bn0 (NMR)	SL3 HAIRPIN FROM THE PACKAGING SIGNAL OF HIV-1		

As one can see, our program can perform calculations for structures with pseudoknots (Table 3). Finally, we present some calculations on the RNAs for which experimental data are still not described in literature. In the PDB and NDB banks we have found a sufficiently large number of hairpin structures, which, in fact, are parts of a single larger structure. The Φ-value profiles for them are quite conservative and almost unchanged even when the length of the hairpin increases (see Table 3). It is evident that such a simple structure as a hairpin has M-shaped Φ-value profiles by our algorithm. It is logical to assume that the terminal nucleotides or long-range contacts are included in the folding nucleus of the hairpin, while the nucleotides of the loops do not form contacts at all, so the Φ-values for them are close to zero.

4. Conclusions

The great diversity of RNA biological functions implies that different molecules may employ different folding pathways. It is still unclear what structural elements determine the folding of different RNA types and how a molecule chooses particular kinetic folding pathways. Important progress has been made in the prediction of protein folding nuclei by starting from the known 3D structures of native proteins. This progress is due to the analysis of multidimensional networks of the protein folding-unfolding trajectories done using various algorithms [6,13,59]. All these approaches (applied also to the studies of the folding rates) use different approximations and algorithms, consider only the attractive native interactions (the "Gō model" [60]) to reduce the energy frustrations and heterogeneity of interactions, and model the trade-off between the formation of attractive interactions and the loss of conformational entropy during protein folding. These studies also simulate unfolding of known 3D protein structures rather than their folding, but the unfolding is considered close to the mid-transition point, where folding and unfolding pathways coincide according to the detailed balance principle. The investigations allowed the authors to outline the protein folding nuclei in more detail. Despite the relative simplicity of these models, they give a promising (~50%) correlation with experiment.

It is noteworthy that this correlation (0.5) for protein structures is considerably worse than those typical of prediction of protein folding rates. The first obvious reason is that the predicted Φ-values are restricted to the narrow region of 0–1 with an experimental error of ~±0.1, while the observed folding rates (determined with a relatively small experimental error) belong to the wide range of 10^7 s^{-1}–10^{-4} s^{-1}. A more important reason is that the folding nucleus is not as stable to the action of mutations (and thus to the unavoidable errors in energy estimates used to outline them) as a 3D protein structure, and it would be strange to obtain a perfect prediction of the folding nuclei with the same force fields which are still not able to predict the mutation-stable 3D native structure of a protein.

As concerns RNA structures, there are not enough quantitative data on RNA folding. The calculated Φ-values for tRNA molecules are in agreement with the experimental data, showing that nucleotide residues in the D and T hairpin regions are involved in the tRNA structure last, or more exactly, they are not included in the tRNA folding nucleus [15]. High Φ-values in the anticodon hairpin region show that the folding nucleus of tRNA is localized just in that place [25]. The calculation was done for 103 tRNA structures with the resolution better than 3 Å out of 126 considered tRNA structures from the PDB and NDB databases. Prediction of RNA folding nuclei gives a possibility to take a fresh look on the problems of multiple folding pathways of the RNA molecule and stability of RNA. One of the

advantages of our method is that it allows us to make prediction of the folding nucleus for nontrivial RNA motifs, such as pseudoknots. The algorithm for prediction of folding nuclei in proteins has been successfully applied to calculations of folding nuclei in tRNA molecules. We are the first who reported the Φ-values and location of the transition states for tertiary folding of various RNA domains. We are the first who prepared the server for calculation the folding nucleus for RNA and protein [61].

Acknowledgments

The authors are grateful to the reviewers and Editor for their valuable comments and Anna V. Glyakina, Dmitry N. Ivankov, and Elena I. Leonova for their help in the preparations of the figures. This work was supported by the Russian Science Foundation (Grant Number 14-14-00536).

Author Contributions

Conceived and designed the experiments: Oxana V. Galzitskaya and Leonid B. Pereyaslavets. Performed the experiments: Oxana V. Galzitskaya and Leonid B. Pereyaslavets. Analyzed the data: Oxana V. Galzitskaya and Leonid B. Pereyaslavets. Wrote the paper: Oxana V. Galzitskaya. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Woodson, S.A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **2010**, *39*, 61–77.
2. Lee, M.-K.; Gal, M.; Frydman, L.; Varani, G. Real-time multidimensional NMR follows RNA folding with second resolution. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9192–9197.
3. Sali, A.; Shakhnovich, E.; Karplus, M. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* **1994**, *235*, 1614–1636.
4. Socci, N.D.; Onuchic, J.N. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys.* **1995**, *103*, 4732–4744.
5. Galzitskaya, O.V.; Finkelstein, A.V. Folding of chains with random and edited sequences: Similarities and differences. *Protein Eng.* **1995**, *8*, 883–892.
6. Galzitskaya, O.V.; Finkelstein, A.V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11299–11304.
7. Matouschek, A.; Kellis, J.T.; Serrano, L.; Bycroft, M.; Fersht, A.R. Transient folding intermediates characterized by protein engineering. *Nature* **1990**, *346*, 440–445.
8. Matouschek, A.; Kellis, J.T.; Serrano, L.; Fersht, A.R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989**, *340*, 122–126.
9. Fersht, A.R. Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1525–1529.

10. Fersht, A.R.; Matouschek, A.; Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992**, *224*, 771–782.
11. Fernández-Escamilla, A.M.; Cheung, M.S.; Vega, M.C.; Wilmanns, M.; Onuchic, J.N.; Serrano, L. Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2834–2839.
12. Sato, S.; Fersht, A.R. Searching for multiple folding pathways of a nearly symmetrical protein: Temperature dependent phi-value analysis of the B domain of protein A. *J. Mol. Biol.* **2007**, *372*, 254–267.
13. Alm, E.; Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11305–11310.
14. Muñoz, V.; Eaton, W.A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11311–11316.
15. Maglott, E.J.; Goodwin, J.T.; Glick, G.D. Probing the Structure of an RNA Tertiary Unfolding Transition State. *J. Am. Chem. Soc.* **1999**, *121*, 7461–7462.
16. Silverman, S.K.; Cech, T.R. An early transition state for folding of the P4-P6 RNA domain. *RNA* **2001**, *7*, 161–166.
17. Kim, J.; Shin, J.-S. Probing the transition state for nucleic acid hybridization using phi-value analysis. *Biochemistry* **2010**, *49*, 3420–3426.
18. Hammond, G.S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334–338.
19. Matouschek, A.; Fersht, A.R. Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7814–7818.
20. Fersht, A. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding; W.H. Freeman: New York, NY, USA, 1999.
21. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **1948**, *437*, 55–75.
22. Svergun, D.I.; Feĭgin, L.A.; Taylor, G.W. *Structure Analysis by Small-angle X-ray and Neutron Scattering*; Plenum Press: New York, NY, USA, 1987.
23. Merino, E.J.; Wilkinson, K.A.; Coughlan, J.L.; Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231.
24. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J. Am. Chem. Soc.* **2005**, *127*, 4659–4667.
25. Pereyaslavets, L.B.; Baranov, M.V.; Leonova, E.I.; Galzitskaya, O.V. Prediction of folding nuclei in tRNA molecules. *Biochemistry* **2011**, *76*, 236–244.
26. Pereyaslavets, L.B.; Sokolovsky, I.V.; Galzitskaya, O.V. FoldNucleus: Web server for the prediction of RNA and protein folding nuclei from their 3D structures. *Bioinformatics* **2015**, *31*, 3374–3376.
27. De Gennes, P.G. Statistics of branching and hairpin helices for the dAT copolymer. *Biopolymers* **1968**, *6*, 715–729.

28. Gutin, A.M.; Galzitskaia, O.V. [Helix-coil transition in the simplest model of large native RNA. I. Consideration of only native helices]. *Biofizika* **1993**, *38*, 84–92.
29. Galzitskaia, O.V. [Effect of the energy of distant contacts on the time of finding the native structure for RNA-like heteropolymers]. *Mol. Biol.* **1997**, *31*, 488–491.
30. Galzitskaya, O.V. Geometrical factor and physical reasons for its influence on the kinetic and thermodynamic properties of RNA-like heteropolymers. *Fold. Des.* **1997**, *2*, 193–201.
31. Zuker, M.; Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **1981**, *9*, 133–148.
32. McCaskill, J.S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **1990**, *29*, 1105–1119.
33. Schuster, P.; Fontana, W.; Stadler, P.F.; Hofacker, I.L. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Biol. Sci.* **1994**, *255*, 279–284.
34. Rivas, E.; Eddy, S.R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **1999**, *285*, 2053–2068.
35. Sato, K.; Kato, Y.; Hamada, M.; Akutsu, T.; Asai, K. IPknot: Fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* **2011**, *27*, i85–i93.
36. Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V.V.; Broude, N.E.; Dokholyan, N.V. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* **2008**, *14*, 1164–1173.
37. Denesyuk, N.A.; Thirumalai, D. Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901–4911.
38. Mathews, D.H.; Sabina, J.; Zuker, M.; Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940.
39. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem. FEBS* **1977**, *80*, 319–324.
40. Jacobson, H.; Stockmayer, W.H. Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *J. Chem. Phys.* **1950**, *18*, 1600–1606.
41. Dawson, W.; Yamamoto, K.; Kawai, G. A new entropy model for RNA: Part I. A critique of the standard Jacobson-Stockmayer model applied to multiple cross links. *J. Nucleic Acids Investig.* **2012**, *3*, 3, doi:10.4081/jnai.2012.2650.
42. Finkel'shteĭn, A.V.; Badretdinov, A.I. [Physical reasons for rapid self-organization of a stable spatial protein structure: Solution of the Levinthal paradox]. *Mol. Biol.* **1997**, *31*, 469–477.
43. Caliskan, G.; Hyeon, C.; Perez-Salas, U.; Briber, R.M.; Woodson, S.A.; Thirumalai, D. Persistence length changes dramatically as RNA folds. *Phys. Rev. Lett.* **2005**, *95*, 268303.
44. Dawson, W.; Yamamoto, K.; Shimizu, K.; Kawai, G. A new entropy model for RNA: Part II. Persistence-related entropic contributions to RNA secondary structure free energy calculations. *J. Nucleic Acids Investig.* **2013**, *4*, 2.
45. Shen, N.; Guo, L.; Yang, B.; Jin, Y.; Ding, J. Structure of human tryptophanyl-tRNA synthetase in complex with tRNATrp reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res.* **2006**, *34*, 3246–3258.

46. Eiler, S.; Dock-Bregeon, A.; Moulinier, L.; Thierry, J.C.; Moras, D. Synthesis of aspartyl-tRNA(Asp) in Escherichia coli—A snapshot of the second step. *EMBO J.* **1999**, *18*, 6532–6541.
47. Galzitskaia, O.V. [Sensitivity of the folding pathway to the details of amino acid sequence]. *Mol. Biol.* **2001**, *36*, 386–390.
48. Semrad, K.; Green, R.; Schroeder, R. RNA chaperone activity of large ribosomal subunit proteins from Escherichia coli. *RNA* **2004**, *10*, 1855–1860.
49. Adams, P.L.; Stahley, M.R.; Kosek, A.B.; Wang, J.; Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **2004**, *430*, 45–50.
50. Bartley, L.E.; Zhuang, X.; Das, R.; Chu, S.; Herschlag, D. Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA. *J. Mol. Biol.* **2003**, *328*, 1011–1026.
51. Deras, M.L.; Brenowitz, M.; Ralston, C.Y.; Chance, M.R.; Woodson, S.A. Folding mechanism of the Tetrahymena ribozyme P4-P6 domain. *Biochemistry* **2000**, *39*, 10975–10985.
52. Greenfeld, M.; Solomatin, S.V.; Herschlag, D. Removal of covalent heterogeneity reveals simple folding behavior for P4-P6 RNA. *J. Biol. Chem.* **2011**, *286*, 19872–19879.
53. Orden, A.V.; Jung, J. Review fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers* **2008**, *89*, 1–16.
54. Ma, H.; Proctor, D.J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P.C.; Gruebele, M. Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.* **2006**, *128*, 1523–1530.
55. Sánchez, I.E.; Kiefhaber, T. Origin of unusual phi-values in protein folding: Evidence against specific nucleation sites. *J. Mol. Biol.* **2003**, *334*, 1077–1085.
56. Juneau, K.; Cech, T.R. *In vitro* selection of RNAs with increased tertiary structure stability. *RNA* **1999**, *5*, 1119–1129.
57. Silverman, S.K.; Zheng, M.; Wu, M.; Tinoco, I.; Cech, T.R. Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* **1999**, *5*, 1665–1674.
58. Staple, D.W.; Butcher, S.E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005**, *3*, doi:10.1371/journal.pbio.0030213.
59. Ivankov, D.N.; Finkelstein, A.V. Protein folding as flow across a network of folding-unfolding pathways. 1. The mid-transition case. *J. Phys. Chem. B* **2010**, *114*, 7920–7929.
60. Taketomi, H.; Ueda, Y.; Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **1975**, *7*, 445–459.
61. Pereyaslavets, L.B.; Sokolovsky, I.V.; Galzitskaya, O.V. FoldNucleus: Web server for the prediction of RNA and protein folding nuclei from their 3D structures. *Bioinformatics* **2015**, *31*, 3374–3376.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).