Annu. Rev. Phys. Chem. 2001. 52:751–62

EARLY EVENTS IN RNA FOLDING

D Thirumalai[1,2,*], Namkyung Lee[2], Sarah A Woodson[3], and DK Klimov[2]

[1]*Department of Chemistry and Biochemistry and* [2]*Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742;*
e-mail: thirum@glue.umd.edu; lee@mpip-mainz.mpg.de
[3]*Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218;*
e-mail: swoodson@jhu.edu
[*]*corresponding author*

Key Words Kinetic partitioning mechanism, RNA folding pathways, *Tetrahymena* ribozyme, specific and nonspecific collapse, counterion-condensation

■ **Abstract** We describe a conceptual framework for understanding the way large RNA molecules fold based on the notion that their free-energy landscape is rugged. A key prediction of our theory is that RNA folding can be described by the kinetic partitioning mechanism (KPM). According to KPM a small fraction of molecules folds rapidly to the native state whereas the remaining fraction is kinetically trapped in a low free-energy non-native state. This model provides a unified description of the way RNA and proteins fold. Single-molecule experiments on *Tetrahymena* ribozyme, which directly validate our theory, are analyzed using KPM. We also describe the earliest events that occur on microsecond time scales in RNA folding. These must involve collapse of RNA molecules that are mediated by counterion-condensation. Estimates of time scales for the initial events in RNA folding are provided for the *Tetrahymena* ribozyme.

INTRODUCTION

The discovery that RNA molecules can function as enzymes (1) has lead to intense efforts to determine their three-dimensional structures and the mechanisms by which they fold. Because RNA molecules are engaged in diverse activities such as ligand binding, protein recognition and catalysis, it is reasonable to suggest that in some ways they are more similar to proteins than to their chemical counterpart, DNA. As in proteins, there are two aspects to RNA folding. The first is the prediction of the folded structures starting from the one-dimensional primary nucleotide sequences. The second is the elucidation of the mechanisms by which they fold. In the past 6 years considerable progress has been made in both aspects of the problem. The determination of the three-dimensional structure of a 160-nucleotide subdomain of *Tetrahymena* group I intron (2, 3), which is the most intensely studied model system for large RNA molecules, has provided

considerable insights into the role of discrete metal ion binding sites in directing folding of ribozymes. Similarly, studies from several groups (4–15) have contributed to our understanding of the folding mechanisms of RNA.

A complete understanding of RNA folding and function requires coming to grips with issues that have already been encountered in the study of protein structure and enzymology. To appreciate the common themes that naturally arise in studies of proteins and RNA, it is instructive to compare the properties of their sequence and conformational space. RNA sequences are formed from four nucleotides, whereas proteins are synthesized from 20 naturally occurring amino acids. Despite this difference, it is easy to show that the number of possible sequences for a moderate-sized RNA or a protein is astronomically large. From an evolutionary perspective it is natural to wonder how the dense manifold of sequence space maps onto the sparse structure space. It is believed that the number of distinct folds in proteins is between 1000 and 4000. The tolerance of proteins to multiple mutations and the relatively small number of folds suggest that a given fold may be the native state for a large number of sequences. These arguments have been made precise using simple models for proteins (16). Because estimates of the number of distinct RNA folds are not available, it is not possible to decipher whether a large number of sequences have very similar tertiary folds. For RNA a simple mapping from sequence to structure space may not be straightforward because the restricted chemical diversity of the building blocks (four nucleotides) may lead to a high degree of structural diversity. A recent experiment has shown that a single nucleotide sequence has two functionally different RNA structures (17). Despite the complications in determining RNA structures it is likely that the mapping from the dense sequence space to sparse structure space should be similar in proteins and RNAs.

The conformational space of RNAs and proteins is extremely large because the number of microstates adopted by their building blocks (nucleotide or amino acids) is greater than two. How RNA navigates this large ensemble of unfolded structures in search of the relatively unique native state in finite time may be stated in the form of the Levinthal "paradox" (18). This argument was first made in the context of protein folding. A resolution of this seeming paradox has brought about a deep understanding of the mechanisms by which proteins fold (18). The considerations described above suggest that the folding kinetics of RNA and proteins should have much in common. Using this hypothesis we suggested a conceptual framework for understanding the folding of large RNA molecules (19). The basic prediction of this framework has found experimental support (8, 10, 13).

In this report we present our view of RNA folding kinetics from an energy landscape perspective. Our work reinforces the notion that there are common themes in the folding kinetics of RNA and proteins (5, 19, 20). Despite the expected similarities in the way RNAs and proteins fold there is a fundamental difference, which clearly affects the earliest events in their folding. The driving force for protein folding is the attractive interactions between the hydrophobic residues that cause the initial collapse (21). In RNA each phosphate group is negatively charged,

so the molecule is a strongly charged polyanion. The initial events must involve substantial neutralization of these charges by counterion condensation. Thus, it is not possible to describe the earliest events in RNA folding without invoking the role of counterions. After describing the overall folding mechanisms we provide a tentative picture of the initial events in RNA folding.

The model system for probing the folding kinetics of large RNA is the *Tetrahymena* ribozyme, an autocatalytic group I intron. This system is an excellent laboratory for probing the folding not only of the intact molecule, but also its subdomains. Throughout this article we use results from this model system (or its variants) to illustrate the fundamental concepts of RNA assembly. Needless to say, our conclusions are general and apply to other large RNA molecules as well.

CONCEPTUAL FRAMEWORK: Kinetic Partitioning Mechanism

We developed a conceptual framework for describing the folding of ribozymes summarized in terms of the kinetic partitioning mechanism (KPM). The underpinnings of our framework are built on the premise that the broad outlines of the way proteins and RNA self-assemble should be similar. The kinetic partitioning mechanism can be understood in terms of the rugged free-energy landscape governing RNA folding. The ruggedness arises for two reasons: (*a*) There are competing interactions in RNA molecules. Attractive interactions (such as base stacking and van der Waals forces) tend to collapse RNA, whereas highly polar and charged moieties are better accommodated by extended structures. Because it is not possible to satisfy all the interactions at every site simultaneously, the molecule is energetically frustrated. (*b*) Due to chain connectivity, there is a high probability of forming local structures to minimize energetic frustration. However, such local structures would be incompatible with the global fold, which requires the formation of specific tertiary contacts. The conflict between local requirements and global considerations, which arises because of the polymeric nature of biomolecules, leads to "topological frustration." If frustration effects are minimized, as is possible for small, well-designed sequences such as tRNA, the free-energy landscape can be relatively smooth. For large RNAs there are many ways of assembling energetically favorable local structures, only one (or few) of which is (are) compatible with the global fold. Some of the incorrectly folded structures can have many aspects in common with the native state. The presence of these low free-energy, native-like misfolded structures could serve as kinetic traps that slow down the folding process.

A schematic sketch of the free-energy surface is shown in Figure 1, from which the basic notions of the KPM can be obtained. Imagine the process, by which an ensemble of unfolded molecules (**U**) begins to navigate the rugged free-energy landscape in search of the native state. The **U** state is a collection of structures that rapidly interconvert among each other. However, at the molecular level, the unfolded structures have considerable heterogeneity. As a result, under

folding conditions, there is a fraction of molecules, Φ, whose conformations map directly onto the native structure. These molecules fold rapidly without populating any discernible intermediates. The remaining fraction, $1-\Phi$, is trapped by one of the deep local metastable minima (Figure 1). Since subsequent rearrangements require activation transitions over free-energy barriers, folding of this set of molecules is slow. Due to the multivalley structure of the free-energy surface (Figure 1) the ensemble of initially unfolded molecules partitions into fast folders (Φ being their fraction) and slow folders.

A consequence of KPM is that the time dependence of a fraction of native RNA should be given by

$$f_N(t) = \Phi_{max} - \Phi e^{-k_{fast}t} - \sum_i A_i e^{-k_i t}, \qquad 1.$$

where k_{fast} is the rate for the fast process, k_i is the rate for transitions from the collection of distinct intermediates **I** to **N**, and A_i is the corresponding amplitude satisfying the condition $\sum_i A_i = \Phi_{max} - \Phi$. If the folding reaction goes to completion, as is assumed in the discussions to follow, $\Phi_{max} = 1$.

The partition factor Φ determines whether a high yield of **N** is produced rapidly. The precise value of Φ (which is proportional to the volume of the native basin of attraction (NBA) in the multidimensional free-energy surface) is determined by the structure of the energy landscape. Thus, Φ depends on the sequence and external conditions, such as temperature, ionic strength etc. We expect Φ to depend on mutations as well. This has been found in a recent experiment (7), which shows that a single point mutation (U273A) in *Tetrahymena* ribozyme drastically increases Φ to 80% in the mutant from 2% in the wild type.

Theoretical arguments have shown that the timescale for the fast folding for RNA with M nucleotides $\tau_{fast} (\sim k_{fast}^{-1})$ scales as (23)

$$\tau_{fast} \simeq \tau_0 M^{\omega} \qquad 2.$$

($\omega \approx 3.8$). The prefactor τ_0, which depends on such factors as viscosity, ionic strength, and monomer diffusion constant, is roughly 10^{-9} s. This formula allows us to estimate τ_{fast} for the fast process within typically an order of magnitude.

KINETICS OF INDEPENDENTLY FOLDING SUBDOMAINS OF *Tetrahymena* RIBOZYME

The framework leading to KPM is also applicable to describe folding of small RNAs (tRNA, for example) as well as independently folding subdomains of the *Tetrahymena* ribozyme. These sequences are expected to fold via two-state kinetics, so that $f_N(t) \simeq 1 - e^{-k_{fast}t}$. Both the P4–P6 domain (Figure 2) as well as the three helix P5abc of the *Tetrahymena* group I ribozyme can fold independently. The folding pathways of P5abc depend on the concentration of Mg^{2+}. At

low Mg^{2+} concentrations NMR studies showed that the native secondary structure is not stable but becomes so during the tertiary folding at elevated Mg^{2+} concentrations (24). The implication is that at low cation concentrations alternative (misfolded) secondary structures are stabilized. This was noted some time ago in the context of tRNA folding (25). The NMR experiments provide a structural basis for the origin of misfolding at the secondary structure level. At high Mg^{2+} concentrations the native secondary structure is "captured" by tertiary interactions. It has been suggested that the acquisition of the tertiary structure, which forces the formation of native secondary structure in P5abc, occurs by a nucleation-collapse mechanism. The folding nucleus is a substructure, containing discretely bound Mg^{2+} forming the magnesium ion core as well as tertiary interactions involving an A-rich bulge of P5a (26). The 56-nucleotide isolated P5abc is predicted to fold with a time constant of about 20 ms (26), which is consistent with the experimental estimate (24, 27) of 20–35 ms.

The rapid formation of P4–P6 subdomain in the intact *Tetrahymena* ribozyme was established using the time-resolved hydroxyl radical footprinting method. More recently, the same method was used to measure the folding rate of the isolated 160 nucleotides P4–P6 domain in the presence of excess monovalent cations Na^+ (27). The folding time is found to be $\sim$500 ms in the absence of Na^+, and it decreases to 15 ms when 50–100 mM salt is added. This is in a good agreement with time-dependent changes in the fluorescence of a covalently attached pyrene (28). Assuming two-state kinetics our estimate (Equation 2) for folding time for P4–P6 domain (in the absence of Na^+) is $\tau_F \approx \tau_0(l_p)M^\omega \approx 600$ ms. According to this estimate changes in τ_F arise because of the dependence of the prefactor τ_0 on the effective persistence length l_p. For a polyelectrolyte, l_p decreases as C, the concentration of Na^+, increases. Because $\tau_0(l_p) \sim l_p$ we expect τ_F to decrease as C is increased, which is in qualitative accord with experiments (27).

Tetrahymena RIBOZYME FOLDS BY THE KPM

The folding pathways of the *Tetrahymena* ribozyme were first probed by Zarrinkar & Williamson (4) using oligonucleotide hybridization. This ribozyme consists of two major subdomains containing paired (P) regions P4–P6 and P3–P7. As discussed above the stable independently folding subdomain P4–P6 folds in about a second, whereas P3–P7 forms on a much longer timescale (4, 6). The hydroxyl radical footprinting measurements revealed that the peripheral domains (P2, P2.1, and P9.1), which make long-range tertiary interactions, form on an intermediate timescale (6). Thus, under typical experimental conditions (adequate Mg^{2+} and no additional cations), the rate-determining step in the folding of this ribozyme involves proper docking of the P3–P7 domain against the P4–P6 subdomain. Using the kinetics of oligonucleotide hybridization two discrete intermediates were identified. One of them is $\mathbf{I}_1$ (P4–P6 folded) and the other $\mathbf{I}_2$, in which both the major subdomains are nearly formed. According to this hierarchical folding model

the **U** state goes through a series of steps (some of which require Mg^{2+}) en route to the native state **N**.

The hierarchical mechanism, resulting in a sequential folding model for RNA, is only valid if Φ is strictly zero. The free-energy landscape perspective predicts that even for a highly frustrated molecule there is a finite probability, however small, that RNA can fold rapidly to the native state. Thus, there must exist a direct pathway from **U** to **N** so that in general, folding of large RNA must involve parallel pathways. Theoretical estimates for the folding timescale for the fast process for $M = 400-650$ is much less than 30 s, making a direct determination of Φ difficult by gel electrophoresis or oligonucleotide hybridization. An estimate of Φ was made by Pan et al (8), who used native polyacrylamide gel containing Mg^{2+} to monitor the time dependence of native state accumulation $f_N(t)$ for $t \geq 1$ min. Accurate double exponential fits of the data for $t > 1$ min showed that f_N $(t = 1\,\text{min})$ did not coincide with the experimentally measured value. Because of conservation of number of RNA molecules the difference yielded $\Phi \approx 0.08$ for the precursor RNA containing *Tetrahymena* ribozyme (cf Equation 1). Thus, a small fraction (8%) of the molecules reaches the native conformation on timescales less than the earliest time accessible in the gel assay for the native state. This study provided the first experimental demonstration that folding of large RNA can be described in terms of KPM. Subsequent studies have given additional support for our framework (29).

DETERMINATION OF Φ BY SINGLE MOLECULE FLUORESCENCE MEASUREMENTS

The clearest and the most direct evidence that RNA folds by parallel pathways with nonzero Φ comes from the recent single molecule studies of Zhuang et al (29). By attaching fluorescent dyes to the 3′ and 5′ ends of the *Tetrahymena* ribozyme, fluorescent resonance energy transfer (FRET) at the single molecule level could be detected. The time-dependent changes in the conformations of the molecule can be probed by measuring the dynamic changes in the FRET signal as a function of the folding conditions. The normalized FRET value in the **U** state is much less than in the **N** state so that time-dependent changes in the FRET signal can be used as a diagnostic for folding. Zhuang et al prepared an initial state, in which RNA is folded, by equilibrating a buffer containing excess Mg^{2+}. At a specified time the solution was diluted and the system was allowed to equilibrate for about 30 s, which is long enough for RNA to fully unfold. This was followed by addition of a buffer containing Mg^{2+}, which subjected the RNA to folding conditions. The method of alternating between folding and unfolding conditions by changing the Mg^{2+} concentration is similar to the NMR H/D exchange method that is used to probe folding kinetics of proteins. Because the mixing time after adding the buffer containing excess Mg^{2+} is much less than τ_{fast}, folding is not

due to nonequilibrium effects (such as shear forces, for example). Had τ_{fast} been on the order of milliseconds or less (as is the case in P5abc, for example), then nonequilibrium effects could affect the folding process.

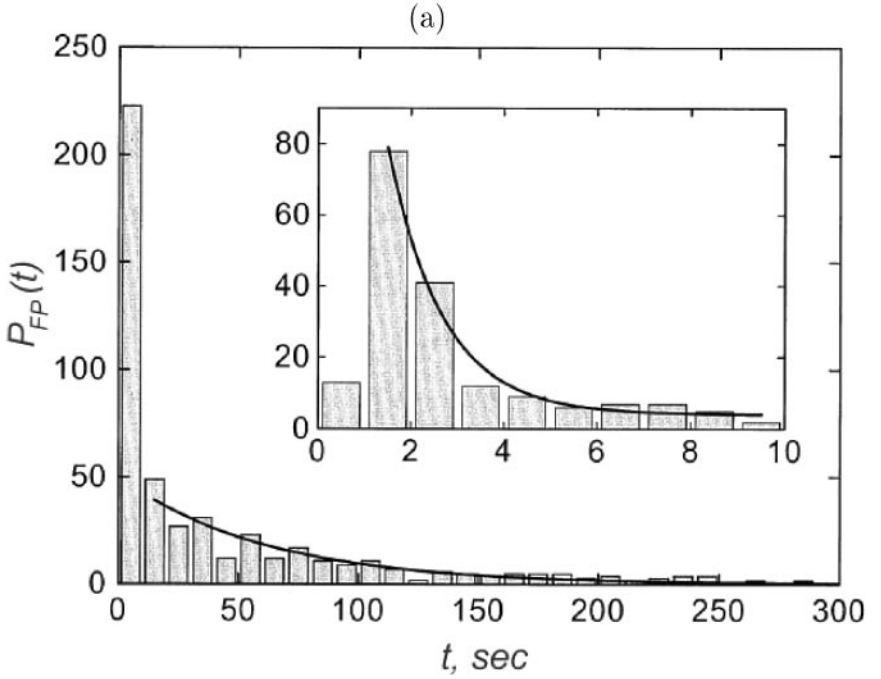
By monitoring the time required for reaching the FRET corresponding to the **N** state the first passage time τ_{1i} for molecule i can be directly measured. Zhuang et al (29) obtained a distribution of first passage times $P_{FP}(t)$ (Figure 3*a*) by measuring τ_{i1} for several hundred molecules from which the fraction of unfolded molecules, $P_u(t)$, can be obtained using

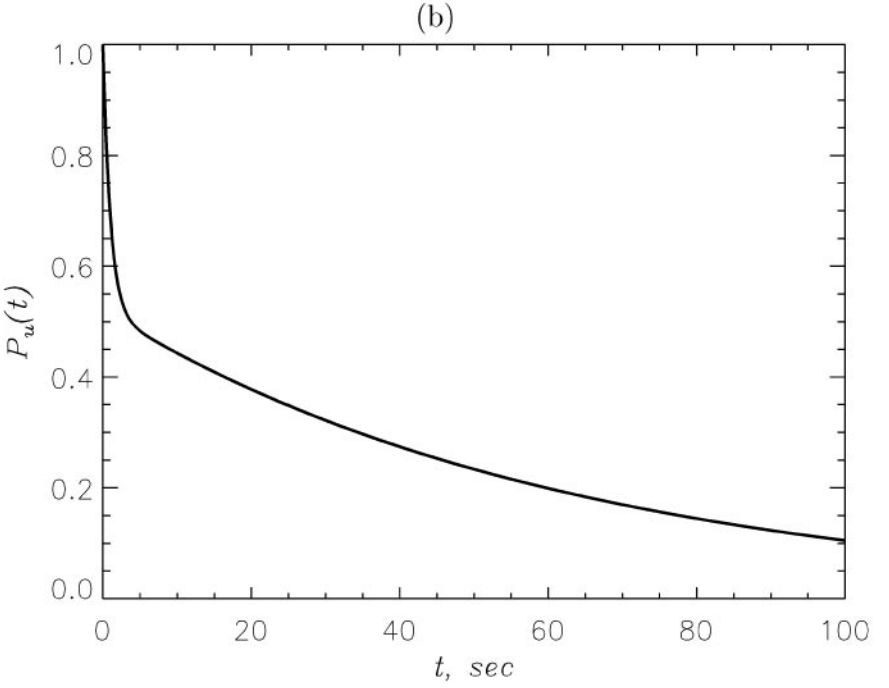
$$P_u(t) = 1 - \int_0^t P_{FP}(s)ds \qquad 3.$$

On the timescale of the single-molecule experiments the ribozyme (29) folds with two distinct timescales. If the KPM mechanism is valid, we expect $P_u(t)$ to be a sum of at least two exponentials. The time constant for the fastest process can be estimated using Equation 2, which for the 400-nucleotide ribozyme is $\tau_{fast} \approx 7$ s. This is in approximate agreement with the experimental value of 1 s, which is obtained by analyzing the data in (Figure 3*a*) for $t < 10$ s. From the fits to $P_u(t)$ (Figure 3*b*) we find that the amplitude of the fastest phase, which is equal to Φ, is the same as that of the slower phase, whose rate constant is $k_1 \simeq 0.016\,\mathrm{s}^{-1}$. Because only 12% of the molecules reach **N** under the conditions of folding, we conclude that $\Phi \approx 0.06$ or 6% of the molecules reach the native state by the fast process. It is noteworthy that the value of Φ is similar for L-21 (29) and for the preRNA (8), which suggests a similarity in the shape of the energy landscape for large RNA molecules.

INITIAL EVENTS IN RNA FOLDING: Overcoming the Electrostatic Repulsion by Counterion-Mediated Collapse

So far we have described events in RNA folding that take place on timescales exceeding at least 1 s. Important folding events that subsequently guide native structure formation occur on subsecond timescales. As in proteins, the earliest events in RNA folding must involve chain compaction. To fold, RNA must overcome the large electrostatic repulsion arising from the Coulomb interactions between the negatively charged phosphate groups. We can estimate the energy due to Coulomb repulsion at temperatures $T > T_m$, the melting temperature. At such temperatures the translational entropy of the counterions is large, and hence they are apparently only weakly attracted to the polyanion. In this case RNA adopts a relatively extended conformation so that the radius of gyration $R_g \sim bM^{\nu}$ (with $\nu \approx 1$), where b is the Kuhn length. The electrostatic repulsion is $E_R \approx \frac{(Me)^2}{\varepsilon R_g} \approx Mk_BT(\frac{l_B}{b})$, where ε is the dielectric constant and $l_B = \frac{e^2}{\varepsilon k_B T}$ is the Bjerrum length. In water, $l_B \simeq 7$Å. Since b is typically less than l_B, we find that $\frac{E_R}{k_BT} \gg 1$ even for relatively small values of M.


(a)
250
200
150
100
50
0
$P_{FP}(t)$
0
50
100
150
200
250
300
t, sec
80
60
40
20
0
0
2
4
6
8
10


(b)
1.0
0.8
0.6
0.4
0.2
0.0
$P_u(t)$
0
20
40
60
80
100
t, sec

Under folding conditions ($T < T_m$ and a buffer solution containing excess Mg_2^+) the electrostatic interactions are sufficiently softened, because the effective charge on the RNA is drastically reduced from *Me*. This happens, just as in polyelectrolytes, by counterion (or Manning) condensation (30). At $T < T_m$ the counterions can condense onto the negatively charged phosphate groups of RNA because the gain in the binding energy exceeds $T\,S_{trans}$, where S_{trans} is the translational entropy of the counterions. When a sufficiently large fraction of the counterions condenses, the effective charge per phosphate group $q \ll e$. Upon counterion condensation, which typically occurs on a diffusion limited timescale (31), RNA adopts a compact conformation. The extent of collapse is determined by a balance between the renormalized Coulomb repulsion and attractive forces (stacking interactions, van der Waals attraction, etc). The size of the counterion-mediated collapsed state $R_g \approx l_p M^{\frac{1}{3}}$. Small-angle X-ray scattering measurements in RNA seem to suggest that $l_p \approx 7$ Å (32) so that the radius of gyration of the collapsed structure of a 657-nucleotide pre-RNA would be about 60 Å.

The nature of the collapsed states as well as the rates of their formation depend on the pathway. Consider the direct pathway, by which a small fraction Φ of molecules reaches the native conformations. We have proposed that along this pathway the processes of collapse and folding are almost simultaneous (23). Here the counterion-mediated collapse is specific, i.e. in the compact phase $\frac{Q_N}{Q_{NN}} > 1$, where Q_N and Q_{NN} are the number of native and nonnative contacts, respectively. Because the number of all topologically allowed nonnative contacts far exceeds Q_N, it follows that the structures formed by specific collapse are native-like. The probability of forming such native-like collapsed structures becomes smaller as RNA becomes larger, which explains the very small values of Φ observed for the *Tetrahymena* folding. The fast folding pathway occurs by the following steps:

$$\mathbf{U} \rightarrow \mathbf{I_S} \rightarrow \mathbf{N}, \qquad 4.$$

where $\mathbf{I_S}$ represents a collection of native-like collapsed structures. The requirement that these structures are native-like implies that the transition from **U** to $\mathbf{I_S}$ may be thermodynamically first order. The rapid conversion of $\mathbf{I_S}$ to **N** also suggests that $\tau_{SC}/\tau_{fast} \leq O(10)$. For the *Tetrahymena* ribozyme we conjecture that the timescale on which $\mathbf{I_{SC}}$ forms is on the order of about 0.1 s.

The majority of molecules in the wild-type *Tetrahymena* ribozyme reach the native state through pathways along which one or two discrete intermediates are

←

Figure 3 Distribution of the first passage times $P_{FP}(t)$ is plotted in (*a*) for *Tetrahymena* ribozyme folding studied by single-molecule experiments (29). Each bin gives the number of RNA molecules that reach the native state within a particular time. The inset shows the blowup of $P_{FP}(t)$ distribution on a small timescale. The time dependence of the fraction of unfolded RNAs $P_u(t)$ obtained from $P_{FP}(t)$ (see text for details) is presented in (*b*). The timescales of the fast and slow pathways are approximately 1 s and 63 s, respectively. The amplitudes of both phases are almost equal.

populated (4). Along these pathways the initial step involves a nonspecific collapse of RNA. This step is expected to be very similar to the collapse of strongly charged polyelectrolytes. With this analogy we propose (31) that the timescale for forming the nonspecific collapse structures should be $\tau_{NSC} \approx \tau_0(l_B, z, S)M^{\alpha}$ (with $\alpha \approx 1$), where z is the valence of counterions. The arguments of τ_0 include the shape S of the counterions. A precise estimate of τ_0 is difficult to make, but we expect it to be between $(10^{-7} - 10^{-6})$s, which implies that $\tau_{NSC} \approx 40$ –400 μs for the ribozyme. The slower pathways can be represented as

$$\mathbf{U} \rightarrow \mathbf{I_{NS}} \rightarrow \{\mathbf{I_i}\} \rightarrow \mathbf{N}, \qquad 5.$$

where $\{\mathbf{I_i}\}$ represents a collection of discrete intermediates. It is possible that some of the intermediates convert to **N** only on timescales exceeding hours. Regardless of the timescales in the formation of the intermediates, it is clear that in all pathways the formation of collapsed structures facilitated by Manning condensation, has to be the earliest event in RNA folding.

CONCLUSIONS

By combining the arguments outlined in this article we summarize the folding pathways in large RNA in Figure 4). A hallmark of this scheme is the demonstration (theoretically and experimentally) that a small fraction of the pool of unfolded molecules folds rapidly by a process involving specific collapse. Only the outlines of the predictions sketched in Figure 4 have been validated. Nevertheless, the conceptual framework allows us to formulate precise questions. A few of these are listed below.

(a) For two-state folders (P5abc and P4–P6 for example) with $\Phi \approx 1$ the structures of the transition states connecting **U** and **N** are not known. The importance of polyelectrolyte interactions in RNA has been used to suggest that the transition states in small RNA molecules should be structurally homogeneous (19). A recent experiment supports this prediction (33). However, additional work is required to probe the TS structures in small RNA molecules. The nature of the transition states connecting $\mathbf{I_i}$ and **N** is to a large extent unknown. RNA engineering experiments will be necessary to provide the structure of these elusive states.

(b) The KPM predicts that point mutations can alter the value of Φ. In the mutant U273A about 80% of the molecules fold in fewer than 30 s. This shows that profound changes in Φ can be achieved by a simple point mutation. Further experiments are needed to understand quantitatively the way Φ depends on the sequence and external conditions.

(c) The prediction that the earliest event in RNA folding must involve the collapse, following counterion condensation, suggests that the pathways,

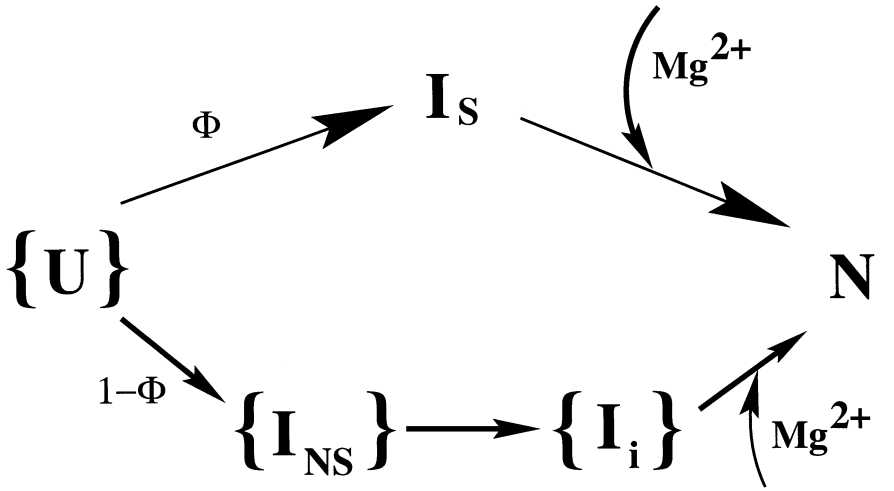


Figure 4 The sketch of the folding pathways for RNA. The fast (upper) folding pathway includes the formation of native-like collapsed states $\mathbf{I_S}$, which rapidly convert into the native state **N**. The fraction of RNA molecules that fold via this pathway is Φ. The lower track (followed by the fraction $1 - \Phi$ of RNAs) represents the slower pathways that involve several stages. First, nonspecific collapse species $\mathbf{I_S}$ form that later convert into a collection of discrete intermediates $\{\mathbf{I_i}\}$. The transition from $\{\mathbf{I_i}\}$ to the native state is slow and represents the rate-limiting step in the slow pathway. The steps requiring binding of magnesium ions Mg^{2+} are also indicated. There could be bifurcations in the pathways leading from $\{\mathbf{I_i}\}$ to **N**. The degree of heterogeneity in the folding pathways depends on the sequence and external conditions.

mechanisms, and timescales must depend on the valence and shape of the counterions. Thus, a complete understanding of RNA folding requires a systematic study of RNA folding kinetics by changing the characteristics of counterions. Because the initial counterion-mediated events occur rapidly, the folding pathways of RNA can only be fully mapped out using fast-folding experiments. Finally, our scheme also suggests that the folding pathways must be heterogeneous. A quantitative validation will require experiments at the single molecule level.

LITERATURE CITED

1. Cech TR. 1993. In *The* RNA *World*, ed. RF Gesteland, JF Atkins, pp. 239–69. Plainview, NY: Cold Spring Harbor Lab. Press
2. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, et al. 1996. *Science* 277:1178–85
3. Batey RT, Rambo RP, Doudna JA. 1999. *Angew Chem. Int. Ed. Eng.* 38:2337–43

4. Zarrinkar PP, Williamson JR. 1994. *Science* 265:918–24
5. Zarrinkar PP, Williamson JR. 1996. *Nat. Struct. Biol.* 3:432–38
6. Scalvi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA. 1998. *Science* 279:1940–43
7. Pan J, Deras ML, Woodson SA. 2000. *J. Mol. Biol.* 296:133–44
8. Pan J, Thirumalai D, Woodson SA. 1997. *J. Mol. Biol.* 273:7–13
9. Treiber DK, Rook MS, Williamson JR. 1998. *Science* 279:1943–46
10. Rook MS, Treiber DK, Williamson JR. 1998. *J. Mol. Biol.* 281:609–20
11. Pan J, Woodson SA. 1998. *J. Mol. Biol.* 280:597–609
12. Sosnick TR, Pan T. 1997. *Nat. Struct. Biol.* 4:931–38
13. Pan T, Fang XW, Sosnick TR. 1999. *J. Mol. Biol.* 286:721–31
14. Treiber DK, Williamson JR. 1999. *Curr. Opin. Struct. Biol.* 9:339–45
15. Thirumalai D, Woodson SA. 2000. *RNA* 6:790–94
16. Thirumalai D, Klimov DK. 2000. In *Stochastic Dynamics and Pattern Formation in Biological and Complex Systems*, ed. S Kim, KJ Lee, W Sung, pp. 95–111. Melville, NY: Am. Inst. Phys.
17. Schultes EA, Bartels DP. 2000. *Science* 289:448–52
18. Dill KA, Chan HS. 1997. *Nat. Struct. Biol.* 4:10–19
19. Thirumalai D, Woodson SA. 1996. *Acc. Chem. Res.* 29:433–39
20. Thirumalai D, Klimov DK, Woodson SA. 1997. *Theor. Chem. Acc.* 96:14–22
21. Dill KA. 1990. *Biochemistry* 29:7133–35
22. Deleted in proof
23. Thirumalai D. 1995. *J. Phys. I* 5:1457–67
24. Wu M, Tinoco I. 1998. *Proc. Natl. Acad. Sci. USA* 95:11555–60
25. Fresco JR, Adams A, Ascione R, Henley D, Linahl T. 1966. *Cold Spring Harbor Symp. Quant. Biol.* 31:527–37
26. Thirumalai D. 1998. *Proc. Natl. Acad. Sci. USA* 95:11506–8
27. Deras ML, Brenowitz M, Ralston M, Chance MR, Woodson SA. 2000. *Biochemistry* 35:10975–85
28. Silverman SK, Deras ML, Woodson SA, Scaringe SA, Cech TR. 2000. *Biochemistry* 39:12465–75
29. Zhuang X, Bartley LE, Babcock AP, Russell R, Ha T, et al. 2000. *Science* 288:2048–51
30. Manning GS. 1977. *Biophys. Chem.* 7: 189–92
31. Lee N, Thirumalai D. 2000. *J. Chem. Phys.* 113:5126–29
32. Russell R, Millett IS, Doniach S, Herschlag D. 2000. *Nat. Struct. Biol.* 7:367–70
33. Maglott EJ, Goodwin JJ, Glick GD. 1999. *J. Am. Chem. Soc.* 7461–62

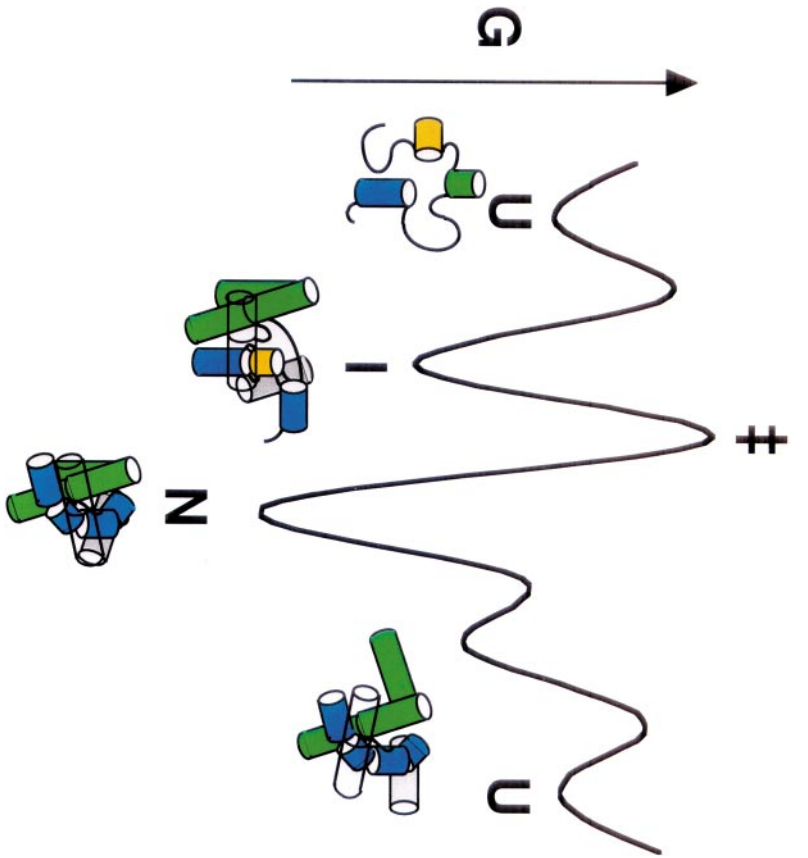


Figure 1 Schematic sketch of the rugged free energy landscape for folding of the *Tetrahymena* ribozyme. The unfolded state is represented as an ensemble of structures in the absence of Mg^{2+}. A small fraction Φ of **U**, whose structures map onto the native state **N**, folds rapidly. The remaining population folds via intermediates (**I**) that are stabilized by native interactions in the P4—P6 domain (*green cylinders*) and misfolded structures in P3 (*yellow*). The slow transition from **I** to **N**, which involves overcoming large activation free energy barrier, requires partial or complete unfolding of the RNA (8).

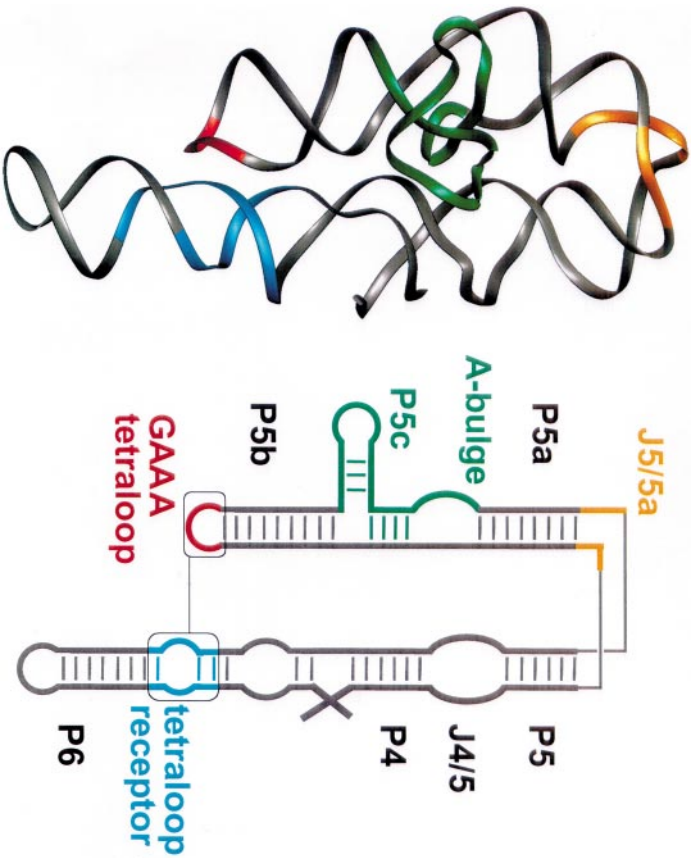


Figure 2 The figure on the right shows a schematic secondary structure of the P4—P6 domain of the *Tetrahymena* ribozyme. The base-paired (P) stems are labeled. On the left a ribbon diagram of the crystal structure of the P4—P6 domain, in the same orientation as the secondary structure, is shown. Comparison of the two gives the locations of the tertiary interactions between the base-paired stems.