Exploring the folding landscape of a structured RNA

Rick Russell*†, Xiaowei Zhuang*‡, Hazen P. Babcock*‡, Ian S. Millett§, Sebastian Doniach‡¶, Steven Chu‡‖, and Daniel Herschlag†‖

Departments of †Biochemistry, ‡Physics, §Chemistry, and ¶Applied Physics, Stanford University, Stanford, CA 94305

Contributed by Steven Chu, November 5, 2001

Structured RNAs achieve their active states by traversing complex, multidimensional energetic landscapes. Here we probe the folding landscape of the *Tetrahymena* ribozyme by using a powerful approach: the folding of single ribozyme molecules is followed beginning from distinct regions of the folding landscape. The experiments, combined with small-angle x-ray scattering results, show that the landscape contains discrete folding pathways. These pathways are separated by large free-energy barriers that prevent interconversion between them, indicating that the pathways lie in deep channels in the folding landscape. Chemical protection and mutagenesis experiments are then used to elucidate the structural features that determine which folding pathway is followed. Strikingly, a specific long-range tertiary contact can either help folding or hinder folding, depending on when it is formed during the process. Together these results provide an unprecedented view of the topology of an RNA folding landscape and the RNA structural features that underlie this multidimensional landscape.

RNAs that adopt specific functional structures face the exacting challenge of attaining these specific structures from a myriad of unfolded conformations. An early and critical step in understanding the folding process is the development of a quantitative kinetic and thermodynamic description. For the *Tetrahymena* group I ribozyme, work over the past decade has established a complex pathway for folding. Intermediates have been identified (1–3), structurally characterized and shown to represent kinetic traps (4–6), and a folding model that unifies many results has been presented (7).

A newer view of folding, first proposed for proteins and explored with lattice models, describes a folding funnel (8–10). In its most rudimentary form, the folding funnel represents a statement that any complex molecular process with many degrees of freedom can be described in terms of a multidimensional energy landscape. For functional macromolecules, kinetic and thermodynamic considerations provide general constraints on the shape of the landscape, necessitating some degree of overall funnel-like shape. That is, folding intermediates leading to the final structure must progress downhill energetically, in general, to allow folding on observed time scales, and the final folded state must sit in a relatively narrow region at the bottom of the energy landscape to ensure that the specific functional state is favored over the ensemble of all other possible conformations (11, 12).

Beyond these requirements, there are many models and disagreements about the topology and nature of barriers in folding landscapes (9, 10, 13, 14), but little experimental data that describe actual folding landscapes beyond individual pathways. For example, it is generally held that folding landscapes for RNAs are more rugged than those for proteins (6, 15, 16), and we believe this to be likely from consideration of the molecular properties of RNA (17, 18). Nevertheless, exploration of folding landscapes will be required to directly address the fundamental question of landscape shape and to provide a framework for revealing potential intermediates and the structural features responsible for "choosing" among the multitude of potential folding pathways (19–21).

The visualization of single molecules provides a powerful means of identifying and characterizing multiple folding pathways (22). However, the ability of this or any other approach to broadly explore features of a folding landscape may be limited by conformational biases present in the starting unfolded conformation (23–25). Such biases may be particularly severe for RNA, which typically has considerable secondary structure present in the starting state. Because there is insufficient time during folding for all unfolded and partially folded conformations on the landscape to be explored (12), a bias in the starting state must give a bias in the portion of the landscape explored during folding. Similarly, a change in the bias of the starting state, leading to population of a different portion of the top of the folding funnel, must lead to exploration of a different part of the landscape during folding. The extent to which folding through different portions of the landscape, or sides of the funnel, leads to gross changes in the folding process remains an open question.

We have therefore turned to a protocol that allows us to populate different ensembles of starting states for the *Tetrahymena* ribozyme, and then rapidly change conditions so that folding proceeds in all cases on the same energy landscape. We follow the folding pathways of the ribozyme by using single-molecule fluorescence methods and demonstrate the presence of discrete folding pathways from the different ensembles of starting states. Structural studies and mutagenesis are then carried out to uncover the structural differences between the starting ensembles and the specific structural features that determine which pathways are traversed.

Materials and Methods

Materials. DNA encoding 5′- and 3′-extended ribozymes was prepared by PCR and cloned into the plasmid pT7L-21 (26). Ribozymes were prepared as described (27). Cy3 and Cy5 (Amersham Pharmacia) labeling of RNA oligonucleotides (Dharmacon, Lafayette, CO) was performed as described (22).

Single-Molecule Fluorescence Resonance Energy Transfer (FRET). Experiments were performed essentially as described (22). For FRET experiments the ribozyme was extended by 5 nucleotides at the 5′ end with the sequence GGUUU. This allowed tight binding of an oligonucleotide substrate (S) capable of base-pairing to the 5′-extension and to the internal guide sequence of the ribozyme (S-CCCUCUAAACC), thereby minimizing dissociation of S from the immobilized ribozyme during experiments. These modifications have no detectable effect on the enzymatic activity of the ribozyme (data not shown).

Small-Angle X-Ray Scattering (SAXS) Measurements. Data were collected on beamline 4–2 at the Stanford Synchrotron Radiation Laboratory essentially as described (ref. 28; see *SAXS Measurements*, which is published as supporting information on the PNAS web site, www.pnas.org).

BIOPHYSICS

Abbreviations: DMS, dimethyl sulfate; FRET, fluorescence resonance energy transfer; S, oligonucleotide substrate; SAXS, small angle x-ray scattering.

*R.R., X.Z., and H.P.B. contributed equally to this work.

‖To whom reprint requests may be addressed. E-mail: schu@leland.stanford.edu or herschla@cmgm.stanford.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Dimethyl Sulfate (DMS) Footprinting. Footprinting was performed essentially as described (29). Only residues that reproducibly gave protections of ≥2-fold are shown as giving protections in Fig. 3. A196, A304, and A306 gave increased DMS accessibility in the presence of Na^+, but these enhancements were not reproducibly larger than 2-fold.

Results

Exploring the Folding Landscape for the *Tetrahymena* Ribozyme. We wanted to initiate folding with RNA molecules having different starting structures. In the language of energy landscapes, we wanted to populate different regions within the unfolded portions of the landscape and then determine what effect, if any, these different starting states had on the folding process. To do this we incubated the *Tetrahymena* group I ribozyme with different concentrations of Na^+ ion, with the idea that varying the cation concentration would change the structure, by ameliorating electrostatic repulsion, without allowing folding to occur, as formation of the native state requires divalent cations (30, 31). We then wanted to add divalent metal ions to observe individual RNA molecules that initiate folding from different structures. If, however, the variable Na^+ concentrations were maintained after addition of Mg^{2+}, the intent of the experiment would be defeated; molecules starting from different structures would also be folding under different conditions, undermining direct comparison. It was therefore critical to rapidly adjust the solution to a constant Na^+ concentration simultaneous with divalent metal ion addition. With this protocol, all of the molecules face the identical energy landscape, and differences in folding behavior can provide information about the topology of the energy landscape.

We followed the folding of individual ribozyme molecules with FRET after preincubation with 20–620 mM Na^+; folding was always performed with 5 mM Mg^{2+} and 20 mM Na^+. Individual ribozyme molecules on a glass surface were visualized by fluorescence emission of Cy3 and Cy5 dyes attached to the S of the ribozyme and to the oligonucleotide tether used in immobilization of the ribozyme (22). For each molecule, we recorded the time at which the FRET value increased to 0.9, indicating native state formation [the ribozyme molecules acquire enzymatic activity concomitantly with the FRET value of 0.9 (22)]. We then plotted the number of native ribozyme molecules as a function of time.

Fig. 1*A* shows results from a representative experiment, with folding following preincubation with 170 mM Na^+. As observed with 20 mM Na^+ present both before and after Mg^{2+} addition (22), folding to the native state is well described by the sum of two exponential curves with rate constants of ≈1 s^{-1} and ≈0.02 s^{-1}. A third folding phase, conversion of a long-lived misfolded state to the native state, does not occur significantly on the time scale of these experiments (7, 27). These two folding phases are collectively referred to as "correct" folding, giving a fraction of the population that folded correctly (i.e., avoided the long-lived misfolded form).

Varying the Na^+ concentration in the preincubation from 20 to 620 mM increased the fraction that folded correctly from 0.2 to ≈0.8 (Fig. 1*B*) even though folding was always under the same conditions (5 mM Mg^{2+}, 20 mM Na^+). Also, the fraction of the correct-folders that folded fast (1 s^{-1}) increased from ≈0.45 to ≈0.9. With high Na^+ concentration in the preincubation (>400 mM), nearly the entire population folded rapidly and to the native state. Replacing Na^+ with other monovalent cations (Li^+, K^+, NH_4^+) gave similar changes in folding properties (not shown). The changes in folding properties, despite folding under identical conditions, suggest that the initial Na^+ concentration affects the starting conformation. Then, from these different starting conformations (or, more precisely, ensembles of starting conformations), different portions of the landscape are traversed during folding.

If Na^+ were shifting the ribozyme between only two conformational ensembles, one of which gave slow and largely incorrect folding whereas the other gave fast and correct folding, the dependencies of both fast and correct folding on initial Na^+ concentration would be the same. However, the changes in fast and correct folding give distinct Na^+ concentration dependencies (Fig. 1*B*), suggesting the presence of at least three conformational ensembles. These three conformational ensembles are as follows. (*i*) One, populated with low Na^+ in the preincubation, gives mostly slow and incorrect folding. (*ii*) With intermediate Na^+ in the preincubation, the populated ensemble gives mostly incorrect folding, but the molecules that fold correctly also fold rapidly. (*iii*) With high Na^+ in the preincubation, the populated ensemble gives mostly fast and correct folding.

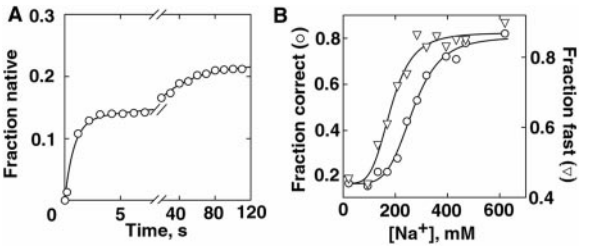
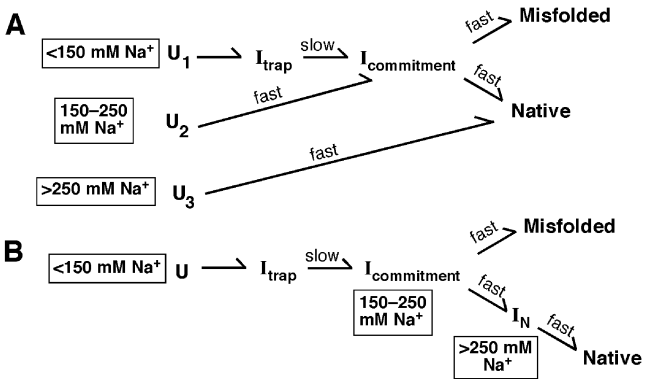


Fig. 1. Exploring the folding landscape. (*A*) Folding was initiated by flowing buffer containing 5 mM Mg^{2+} over the surface-attached ribozyme, replacing the 170 mM Na^+ solution present initially. The fraction of molecules that attained FRET values of ≈0.9, indicating native-state formation, is plotted against folding time. The curve represents a fit by two exponentials with rate constants of 1.21 ± 0.07 s^{-1} and 0.025 ± 0.001 s^{-1}. The combined amplitude of these two folding phases, 0.219 ± 0.003, represents the fraction of ribozyme that folded to the native state without forming the long-lived misfolded state. (*B*) The fraction of ribozyme that folded correctly (○), and the fraction of the correct-folding molecules that folded fast (≈1 s^{-1}, ▽), are plotted against initial Na^+ concentration. Only the correct-folding molecules were analyzed for the rate of folding because the molecules that misfold do not give a detectable change in FRET value after formation of an early intermediate (22). The dependence of fast folding on Na^+ concentration gave a half-maximal value for Na^+ ($K_{1/2}$) of 183 ± 9 mM and a Hill coefficient (*n*) of 4.4 ± 0.9. For correct folding, these values were $K_{1/2}$ = 267 ± 8 mM, n = 5.0 ± 0.7.

Although in principle Na^+ dilution to 20 mM is instantaneous with Mg^{2+} addition and the onset of folding, in practice there is a finite mixing time. Thus, the differences in folding properties could arise if, instead of or in addition to changing the starting conformation, the Na^+ still present only milliseconds after Mg^{2+} addition affected very early folding steps. To test this possibility, we performed an experiment in which Na^+ was washed out with variable time delays before the addition of Mg^{2+}. Although the bulk and RNA-associated Na^+ ions are expected to exchange on the submillisecond time scale (32), the time dependencies for the changes in folding behavior upon removal of Na^+ were on the scale of seconds (Fig. 6, which is published as supporting information on the PNAS web site). These slow transitions provide strong evidence that the changes in folding behavior arise from changes in starting conformation rather than Na^+ effects on very early events in folding. Further, the losses of correct folding and fast folding gave distinct rate constants (0.6 s^{-1} and 0.09 s^{-1}, respectively), confirming the existence of at least two discrete structural transitions induced by Na^+ (i.e., at least three conformational ensembles).

The experiments above build on previous results by Woodson and colleagues (33, 34), who showed that preincubation with Na^+ allows the ribozyme to fold to the native state in the presence of Mg^{2+} within the time required to enter a native gel (≈30 s). Here, we have used the increased time resolution and decreased mixing time in our single-molecule approach to directly demonstrate and follow a fast-folding phase. Further, we have shown that the observed change in folding behavior results



Scheme 1.

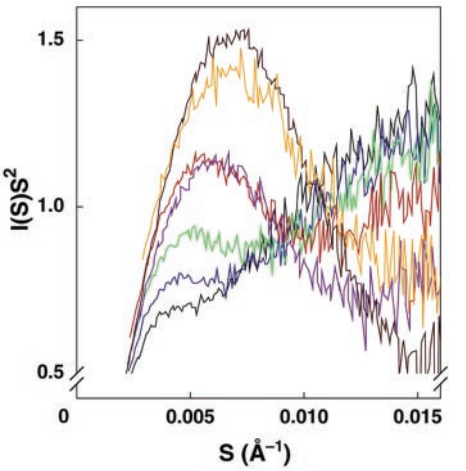


Fig. 2. Structural changes in the starting ensemble monitored by SAXS. Kratky plots are shown for the standard L-21 *Sca*I ribozyme at 25°C in buffer containing 20 mM Na^+ (black), 70 mM Na^+ (blue), 120 mM Na^+ (green), 420 mM Na^+ (red), 820 mM Na^+ (purple), the intermediate I_{trap} (orange, formed at 15°C), and the native ribozyme in buffer containing 15 mM Mg^{2+} (brown). Kratky plots have the form of $I(S)S^2$ against S, where I(S) is scattering intensity and S is related to the scattering angle [$S = 2\sin\theta/\lambda$, where λ is the x-ray wavelength, 1.54 Å, and 2θ is the scattering angle (35)].

from at least two structural transitions; one transition gives fast folding and the other gives correct folding. Below we use these results to further our understanding of the folding landscape, first by providing evidence that the starting states induced by Na^+ give folding along discrete pathways through the folding landscape and then by identifying structural features that determine which pathways are followed.

Determination of Multiple Pathways on the Folding Landscape. Scheme 1 depicts two models that can account for the data described above. [The models also incorporate previous results (7, 22)**.] In Model 1, different Na^+ concentrations cause three different starting states to be differentially populated. These different starting states then fold along discrete pathways through the landscape (Scheme 1*A*). With low concentrations of Na^+ in the preincubation, a significant fraction (0.55) starts folding from U_1. These molecules proceed through the late folding intermediate I_{trap} and then slowly escape this intermediate to partition between the native and long-lived misfolded forms (1, 2, 7). With intermediate Na^+ in the preincubation, folding starts from U_2; I_{trap} is avoided but the partitioning between the native and long-lived misfolded forms is the same as at low Na^+. This identical partitioning suggests the presence of the common folding intermediate $I_{commitment}$ (22). With high Na^+ in the preincubation, U_3 is populated and essentially all of the ribozyme folds fast and correctly; thus, both I_{trap} and $I_{commitment}$ are avoided.

In Model 2, the changes in folding properties arise because increasing Na^+ concentration causes the ribozyme to fold part of the way along a single pathway; i.e., when Mg^{2+} is added the ribozyme is already partially folded (Scheme 1*B*). Folding with $<$150 mM Na^+ begins with U, proceeds through a rate-limiting step escaping from the intermediate I_{trap}, and then, after formation of $I_{commitment}$, partitions between formation of the native and long-lived misfolded states. Folding is faster after preincubation with 150–250 mM Na^+ because an intermediate (e.g., $I_{commitment}$) situated after the rate-limiting step is populated at the start of folding. With still higher Na^+ in the preincubation both I_{trap} and $I_{commitment}$ are avoided so that folding is fast and gives exclusively native ribozyme.

The models in Scheme 1 *A* and *B* were distinguished by performing SAXS on the conformational ensembles induced by Na^+. Model 2 predicts that Na^+ allows the ribozyme to form folding intermediates that are also formed in Mg^{2+} after the rate-limiting step for overall folding (after I_{trap} in Scheme 1*B*). Previous SAXS experiments showed that I_{trap} is nearly as compact as the native state (28) and the simplest expectation is that intermediates after I_{trap} would also be compact (see *Determination of Discrete Pathways by SAXS*, which is published as supporting information on the PNAS web site). As the concentration of Na^+ was increased in the range that gives changes in the folding properties, SAXS profiles showed appearance of a peak in Kratky plots (Fig. 2) (35). This peak indicates that the ribozyme becomes somewhat more compact and globular as the concentration of Na^+ is increased. However, even the highest concentration of Na^+ used (820 mM), sufficient to give essentially all fast and correct folding, gave a Kratky peak much less pronounced than that of the intermediate I_{trap} or the native ribozyme.

These results indicate that the ribozyme is substantially less compact and globular than I_{trap} and the native ribozyme across the range of Na^+ concentrations, providing strong evidence against Model 2. Further, solution radical protection experiments gave no indication of native tertiary structure formation in Na^+ concentrations that give fast and correct folding (ref. 31; M. Brenowitz, personal communication). Thus, the results strongly favor Model 1 (Scheme 1*A*), in which folding starts from different conformational ensembles that depend on the initial Na^+ concentration and proceeds through discrete pathways.

Structural Features That Determine Which Pathway Is Traversed. To identify the structural changes that might be responsible for the different observed folding behavior after preincubation with different concentrations of Na^+, protection of adenosine and cytosine residues from DMS was monitored. As the concentration of Na^+ was increased, residues in P9, P3, and P4-P6 became protected from DMS modification with distinct concentration dependencies (Fig. 3). Interestingly, two sets of protections were observed at Na^+ concentrations in the range that gives changes in folding properties (100–300 mM; Fig. 1*B*). The observed protection of residue C278 in P3 suggests the formation of P3 (3), and the protection of residues in loop L9.1 suggests the formation of P13, a long-range base-pairing contact between residues in L9.1 and L2.1. This interpretation was supported by the finding that a variant ribozyme with a three-nucleotide substitution in L2.1 sequence to disrupt P13 [ΔP13 (36)] does not give protection of L9.1 (data not shown). These results raise the

**In Scheme 1 *A* and *B*, the Na^+ concentration ranges shown indicate which conformational ensemble is most populated. At any given Na^+ concentration, a mixture of these ensembles is expected to exist as shown by the data in Fig. 1*B*.

BIOPHYSICS

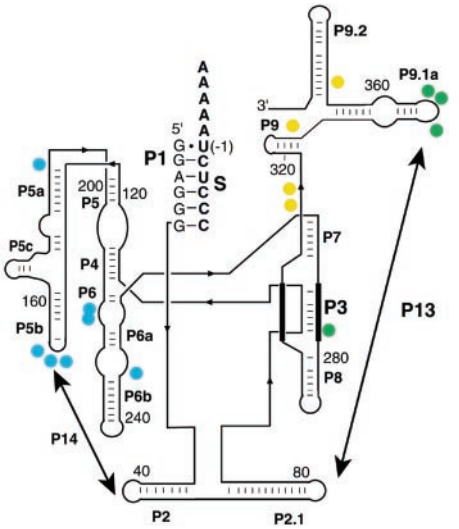


Fig. 3. Structural changes in the starting ensemble monitored by DMS accessibility. Residues that increase in protection from DMS in the presence of NaCl are indicated with filled circles. Residues that gave an Na^+ concentration for half-maximal protection ($K_{1/2}$) < 100 mM are colored yellow (C315, C316, C332, C371), residues that gave $K_{1/2}$ values between 100 mM and 300 mM Na^+ are green (C278, A347, A351, A352), and residues with $K_{1/2}$ > 300 mM Na^+ are blue (A151–153, A196, C216, C217, A246). Results are for the standard L-21 *Sca*I ribozyme in the presence of bound S. The extended version used for single-molecule experiments gave an indistinguishable Na^+ dependence of DMS protection (data not shown), suggesting that the structural changes induced by Na^+ are similar for the two versions of the ribozyme. The strands that pair to form Alt P3, which can replace P3 (3), are shown with thick lines.

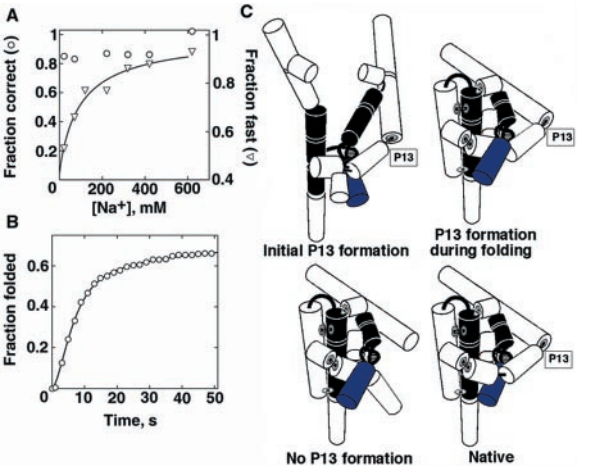


Fig. 4. Structural changes responsible for distinct folding properties. (*A*) Folding of the U273A variant ribozyme, in which P3 is stabilized. The fraction of ribozyme molecules that folded correctly (○) and fast (▽) are plotted against initial Na^+ concentration. A fit by the Hill equation gave values for fast folding of $K_{1/2}$ = 100 ± 50 mM Na^+ and n = 0.9 ± 0.2. (*B*) Folding of the ΔP13 variant (U75A, G76C, C77G), in which P13 is disrupted. The ribozyme also contained the U273A substitution, allowing it to avoid the long-lived misfolded form. An increase in rate at early times (a kinetic lag) was observed, necessitating two rate constants of 0.30 ± 0.04 s^{-1} and 0.29 ± 0.04 s^{-1} to describe the fast phase of folding [620 mM Na^+ initially; similar results were obtained with an initial Na^+ concentration of 20 mM. A small fraction (10%) folded slowly, giving a third rate constant of 0.038 ± 0.005 s^{-1}]. These two larger rate constants are the same within error as that for docking of the P1 duplex (0.23 s^{-1}; unpublished results). As docking is required for efficient FRET (22), the other rate constant of 0.3 s^{-1} is assigned to overall folding. (*C*) Model for the effects of P13 formation on folding (see text). The model reflects folding in the context of preformed P3, shown as two thick lines connected by thin lines. Other secondary structure elements are shown as cylinders, and long-range contacts are shown as thin connecting lines.

possibility that initial formation of P3 and/or P13 in the presence of Na^+ is responsible for the observed fast and correct folding.

P3 formation was particularly appealing as a structural determinant for folding pathways in light of previous studies showing that an alternative local duplex termed Alt P3 replaces P3 in the long-lived misfolded state (3). Further, a variant ribozyme, U273A, in which P3 is stabilized relative to Alt P3, has been observed to fold both fast and correctly (37). These previous results suggested a model in which formation of the correct secondary structure early in folding is critical: if P3 is formed correctly, folding is fast and correct, whereas if Alt P3 is formed early, the ribozyme gets trapped in misfolded species (37, 38).

Our results confirming that correct folding is observed with high initial Na^+ concentration (33) and showing P3 is formed initially under these conditions provides further support for the model that P3 formation is responsible for the correct folding (37, 38). On the other hand, the observation that the conformational transitions that give fast folding and correct folding are distinct (Fig. 1*B*) calls into question the previous conclusion that P3 formation is also responsible for fast folding.

To determine whether early P3 formation is responsible for both fast and correct folding we examined folding of the variant U273A by single-molecule FRET as a function of initial Na^+ concentration. As expected (3, 37), folding of the U273A variant ribozyme was largely correct at all initial Na^+ concentrations (Fig. 4*A*). Further, DMS footprinting revealed that P3 is formed initially in the U273A variant across the entire range of Na^+ concentration (20–820 mM; data not shown). However, with low initial Na^+ concentrations, two rate constants for folding were observed as with the wild type; about half of the variant ribozyme folded slowly (0.012 s^{-1}), approximately the same fraction of slow folding as for the wild type. The concentration of Na^+ required to give fast folding of the U273A variant was similar to that for the wild type. Thus, although initial P3 formation correlates strongly with correct folding, it is not sufficient to give fast folding. It is also suggested that P3 formation is not necessary for fast folding, as fast folding with the wild-type ribozyme occurs at Na^+ concentrations lower than required to give correct folding.

For the U273A variant, as for the wild type, residues in P13 became protected from DMS in the Na^+ concentration range that gave fast folding (not shown). We therefore tested the effect of P13 on fast folding, monitoring the folding of the variant ribozyme in which P13 is disrupted (ΔP13). This substitution was made in the background of the U273A variant to simplify analysis by allowing predominantly correct folding at all Na^+ concentrations. We reasoned that if P13 formation were responsible for fast folding, the ΔP13 variant would give slow folding at all initial concentrations of Na^+ because of the inability to form P13 (36).

Contrary to this prediction, predominantly fast folding was observed after preincubation with both low and high Na^+ concentrations (Fig. 4*B*). This surprising result indicates that the order of tertiary structure formation can be important. That is, if P13 is initially formed, fast folding is observed. Likewise, if P13 cannot form, fast folding is also observed. However, if P13 is not formed initially but has the ability to form during folding, folding is slow for a fraction of the ribozyme.

One general possibility for the origin of these effects is that formation of P13 at the "wrong" time can lead to topological barriers in folding (Fig. 4*C*). According to this model, if P13 is formed initially, as it is with high initial Na^+ concentration for the U273A variant and the wild-type ribozyme, folding is fast because constraints imposed by the preformed P13 give a ribozyme core that is prearranged to fold rapidly. In contrast, if P13 is not formed

initially but has the potential to form, as is the case with low Na^+ concentration for both the U273A and wild-type ribozymes, folding of a large fraction is slow, consistent with a barrier to rearrangement that is created after P13 formation, as suggested (6, 39). Fig. 4*C* shows one specific possibility: after formation of P13, the P8 duplex becomes trapped outside the ring composed of P2, P2.1, and P9. Finally, if P13 is prevented from forming by mutagenesis (ΔP13), barriers created by P13 formation would not be present, allowing the observed fast folding (see *Additional Results from P13 Disruption*, which is published as supporting information on the PNAS web site).

Discussion

When a macromolecule folds to an active structure, it traverses a multidimensional energetic landscape; thus, understanding the energetic topology of the landscape is essential for understanding the folding process for a given molecule. The landscape is a representation of intermediates and their energetic connections with each other and with the native and unfolded states. Beyond this energetic description, structural descriptions of intermediate forms are needed. Further, we want to know what specific parts of the structure are important for defining the different intermediates and the energy barriers for interconversion between them.

To obtain new information about the topology of the landscape for the *Tetrahymena* ribozyme, we used a procedure that allowed us to broadly explore the landscape. Preincubation with different concentrations of Na^+ allows folding to be initiated from different starting structures, i.e., different regions of the folding landscape. To probe the same landscape, folding was then followed under identical conditions in all experiments. With this procedure, we found that the folding properties strongly depend on the region of the landscape from which the molecules start to fold. With increasing concentrations of Na^+ present in the preincubation, the fraction of ribozyme molecules that folded fast increased, as did the fraction that avoided a long-lived misfolded conformation.

Two models were considered to explain this behavior: the different starting structures lead to folding along discrete pathways or the different starting structures instead represent successive intermediates along a single folding pathway. It was shown that the folding properties of the ribozyme depend on starting conditions (33, 34), but the models above were not distinguished. Here, SAXS experiments and our knowledge of the compactness of folding intermediates provided strong evidence for discrete folding pathways. Two discrete folding pathways have also been demonstrated for a circularly permuted variant of RNase P (40).

These results extend our understanding of the folding landscape for the *Tetrahymena* ribozyme (Fig. 5*A*). The landscape contains pathways that lie in deep channels, with barriers, likely with both enthalpic and entropic components, sufficient to prevent interconversion of molecules on the different pathways (see also refs. 41 and 42 for further discussions of discrete pathways). The solution conditions before Mg^{2+}-induced folding determine the starting RNA structures and, thus, which pathways are entered. After incubation in low Na^+ (<150 mM), pathway A (which lies in Channel A in Fig. 5*A*) is predominantly traversed, leading to slow folding by means of I_{trap} and partitioning between N and M, the native and long-lived misfolded states, respectively. Pathways B and C give fast folding by avoiding the intermediate I_{trap}, and pathway C gives correct folding by avoiding the region of the landscape that drains into the long-lived misfolded form M (this corresponds to avoiding $I_{commitment}$ in Scheme 1*A*).

To probe the structural underpinnings of the discrete folding pathways, we used chemical protection to identify structural differences between the starting conformations and then mutagenesis to identify what structural changes are important for

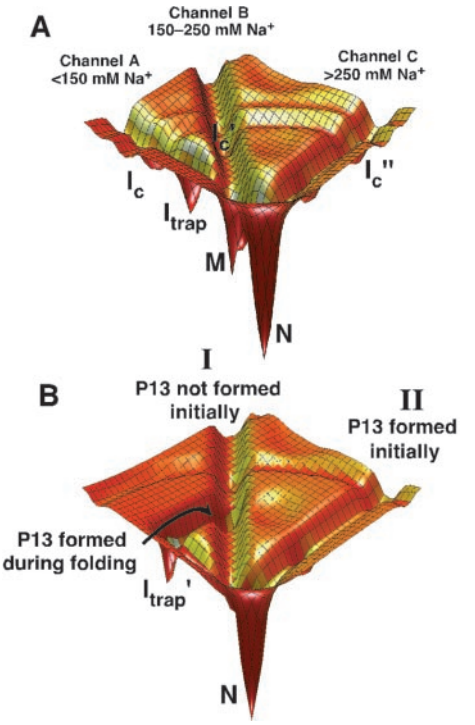


Fig. 5. RNA folding pathways contained in channels. (*A*) A schematic of channels for the wild-type ribozyme. The dependence of folding properties on initial conditions indicates the presence of free energy barriers between the channels, whereas the figure is contoured by internal free energy (9, 10). At least some of these barriers are likely to be present when considering internal free energy because interconversion between the starting states requires base pairs to exchange. Therefore, ''walls'' are shown separating the channels. The simplest model is shown, in which there are three pathways, A–C, that lie in channels A, B, and C, respectively. Pathways A and B predominate at low initial Na^+ concentration, pathway B at intermediate Na^+ concentration, and pathway C at high Na^+ concentration. The fast-folding fraction under low Na^+ conditions could instead arise from an additional pathway that branches from pathway A during folding before the formation of I_{trap} (not shown). I_c is a collapsed intermediate that forms along pathway A (R.R., I.S.M., M. Tate, L. Kwok, B. Nakatani, V. Pande, S. Gruner, S. Mochrie, S.D., D.H., and L. Pollack, unpublished data), and analogous intermediates are postulated along pathways B and C. (*B*) Effects of P13 formation during folding starting with a preformed P3 (see text). The landscape for folding of the U273A ribozyme is depicted for simplicity, in which the long-lived misfolded form is not formed significantly and is therefore not shown.

defining the different pathways. DMS footprinting showed that several regions became protected with increasing concentration of Na^+ and identified the long-range base-pairing contact P3 as a strong candidate for involvement in the changes in folding properties (3, 37).

The involvement of early P3 formation in fast folding and correct folding was tested by following folding of a ribozyme variant in which P3 is stabilized. Folding of this variant showed that formation of P3 in the starting population is responsible for correct folding as suggested (37). On the other hand, we found that P3 formation in the starting population is not necessary or sufficient for fast folding. Thus, multiple structural changes are responsible for determining which pathways are traversed.

Interestingly, when P3 is formed initially in the variant, DMS protection experiments suggest that essentially all of the secondary structure is correctly formed (unpublished results); nevertheless, folding can be slow. This finding suggests that there are substantial energy barriers for RNA folding from preformed secondary structure, such as topological constraints on the

dissolution of fortuitously formed nonnative tertiary interactions and on movements of subdomains relative to each other.

To further probe the structural features that determine which pathway is traversed, we explored the effects of disrupting the long-range contact P13, which was also shown by DMS footprinting to form in the presence of Na^+. P13 formation is involved with fast folding but in a surprising manner: forming the right structure, P13 can either help or hinder folding, depending on when it is formed during the folding process.

Forming P13 during the folding process is deleterious, leading to slow folding. This result is analogous to results showing that native interactions in P5abc can slow fold by stabilizing a kinetically trapped intermediate (4, 5) and can be rationalized by the location of P13 on the exterior of the ribozyme model structure (36). The deleterious effects of P13 could arise because a topological barrier to folding introduced by P13 formation prevents access of other structural elements to the molecule's core (Fig. 4*C*) or because P13 formation brings together regions of the molecule that have yet to establish their native interactions and are thus highly susceptible to formation of fortuitous nonnative interactions.

Remarkably, if P13 is formed early, at the outset of folding, it no longer hinders folding. This difference in the effect of P13 formation could arise because when P13 is formed in the starting state the core elements are preoriented to fold efficiently as depicted in Fig. 4*C*, or it could be because early P13 formation blocks a nonnative contact that otherwise forms during folding and is subsequently stabilized if P13 forms later. Regardless of the molecular origin, this result indicates that the presence of P13 is not alone responsible for and does not necessitate the formation of a kinetic trap. This finding is depicted in the simplified energy landscape in Fig. 5*B*. If P13 is not formed initially, pathway I is followed; after formation of P13, the RNA is shunted to form a trapped species (I_{trap}'). Thus, if P13 is mutated and cannot form, the trap is avoided and folding to N is fast. The avoidance of the trap when P13 is formed at the outset of folding indicates that a distinct folding pathway is followed (Fig. 5*B*, pathway II); the barriers separating pathway II from pathway I are sufficient to prevent significant formation of I_{trap}'.

Multiple kinetic phases are commonly observed for RNA folding (e.g., refs. 3, 27, 40, and 42–46), indicating the existence of multiple folding pathways. Thirumalai and Woodson have proposed that these pathways are separated from the earliest stages of folding (41). Results here provide direct evidence that some pathways are separated from the earliest stages of folding (Fig. 5*A*), and also demonstrate the existence of a slow-folding pathway that is separated by energy barriers only later in folding, after P13 formation (Fig. 5*B*). We showed that there is also a separation of pathways very late in folding that commits the ribozyme to fold to the native state or to a long-lived misfolded state (7). Thus, there are likely to be networks of folding pathways throughout the landscape that branch and reconnect at different points leading to formation of trapped species and the native ribozyme.

Recent work in protein folding has emphasized that the unfolded states can contain residual structure, and this residual structure is suggested to be important for the folding process (24, 47). For RNA it is particularly striking that the starting states are not simply "unfolded" but rather contain substantial structure. Further, results here show that the structural features in the starting states and structural differences between different starting states can have profound effects on folding. Future characterization of these folding pathways will clarify the relationship of these starting structures to intermediates on the folding pathways. Do the starting structures in Na^+ give enhanced folding because they resemble late-folding intermediates, or are they conformations that are largely unfolded but are readily transformed into the late-folding intermediates? The methods used here should be valuable as we continue probing the pathways identified here, and continue exploration to uncover and understand new pathways and features of this landscape.

This research was supported by a Packard Interdisciplinary grant (to D.H. and S.C.), a National Science Foundation grant, and by an Air Force Office of Scientific Research grant (to S.C.). The U.S. Department of Energy and the National Institutes of Health support the Stanford Synchrotron Radiation Laboratory. R.R. and X.Z. were supported by National Institutes of Health postdoctoral fellowships. H.P.B. was supported in part by a Center on Polymer Interfaces and Macromolecular Assemblies grant.

1. Zarrinkar, P. P. & Williamson, J. R. (1994) *Science* **265,** 918–924.
2. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. (1998) *Science* **279,** 1940–1943.
3. Pan, J. & Woodson, S. A. (1998) *J. Mol. Biol.* **280,** 597–609.
4. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. (1998) *Science* **279,** 1943–1946.
5. Rook, M. S., Treiber, D. K. & Williamson, J. R. (1998) *J. Mol. Biol.* **281,** 609–620.
6. Treiber, D. K. & Williamson, J. R. (1999) *Curr. Opin. Struct. Biol.* **9,** 339–345.
7. Russell, R. & Herschlag, D. (2001) *J. Mol. Biol.* **308,** 839–851.
8. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8721–8725.
9. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21,** 167–195.
10. Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4,** 10–19.
11. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. J. (1961) *Proc. Natl. Acad. Sci. USA* **49,** 1309–1314.
12. Levinthal, C. (1969) in *Proceedings of a Meeting Held at Allerton House, Monticello, IL*, eds. Debrunner, P., Tsibris, J. C. M. & Münck, E. (Univ. of Illinois Press, Urbana, IL), pp. 22–24.
13. Baldwin, R. L. & Rose, G. D. (1999) *Trends Biochem. Sci.* **24,** 77–83.
14. Bilsel, O. & Matthews, C. R. (2000) *Adv. Protein Chem.* **53,** 153–207.
15. Woodson, S. A. (2000) *Cell. Mol. Life Sci.* **57,** 796–808.
16. Treiber, D. K. & Williamson, J. R. (2001) *Curr. Opin. Struct. Biol.* **11,** 309–314.
17. Sigler, P. B. (1975) *Annu. Rev. Biophys. Bioeng.* **4,** 477–527.
18. Herschlag, D. (1995) *J. Biol. Chem.* **270,** 20871–20874.
19. Jennings, P. A., Finn, B. E., Jones, B. E. & Matthews, C. R. (1993) *Biochemistry* **32,** 3783–3789.
20. Radford, S. E., Dobson, C. M. & Evans, P. A. (1992) *Nature (London)* **358,** 302–307.
21. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9029–9033.
22. Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D. & Chu, S. (2000) *Science* **288,** 2048–2051.
23. Srinivasan, R., Rose, G. D., Leopold, P. E., Montal, M. & Onuchic, J. N. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14258–14263.
24. Shortle, D. & Ackerman, M. S. (2001) *Science* **293,** 487–489.
25. Murthy, V. L., Srinivasan, R., Draper, D. E. & Rose, G. D. (1999) *J. Mol. Biol.* **291,** 313–327.
26. Zaug, A. J., Grosshans, C. A. & Cech, T. R. (1988) *Biochemistry* **27,** 8924–8931.
27. Russell, R. & Herschlag, D. (1999) *J. Mol. Biol.* **291,** 1155–1167.
28. Russell, R., Millett, I. S., Doniach, S. & Herschlag, D. (2000) *Nat. Struct. Biol.* **7,** 367–370.
29. Inoue, T. & Cech, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 648–652.
30. Latham, J. A. & Cech, T. R. (1989) *Science* **245,** 276–282.
31. Celander, D. W. & Cech, T. R. (1991) *Science* **251,** 401–407.
32. Bleam, M. L., Anderson, C. F. & Record, M. T. (1983) *Biochemistry* **22,** 5418–5425.
33. Heilman-Miller, S. L., Thirumalai, D. & Woodson, S. A. (2001) *J. Mol. Biol.* **306,** 1157–1166.
34. Heilman-Miller, S. L., Pan, J., Thirumalai, D. & Woodson, S. A. (2001) *J. Mol. Biol.* **309,** 57–68.
35. Glatter, O. & Kratky, O. (1982) *Small Angle X-ray Scattering* (Academic, London).
36. Lehnert, V., Jaeger, L., Michel, F. & Westhof, E. (1996) *Chem. Biol.* **3,** 993–1009.
37. Pan, J., Deras, M. L. & Woodson, S. A. (2000) *J. Mol. Biol.* **296,** 133–144.
38. Thirumalai, D. & Woodson, S. A. (2000) *RNA* **6,** 790–794.
39. Pan, J. & Woodson, S. A. (1999) *J. Mol. Biol.* **294,** 955–965.
40. Pan, T., Fang, X. & Sosnick, T. (1999) *J. Mol. Biol.* **286,** 721–731.
41. Thirumalai, D. & Woodson, S. A. (1996) *Acc. Chem. Res.* **29,** 433–439.
42. Pan, J., Thirumalai, D. & Woodson, S. A. (1997) *J. Mol. Biol.* **273,** 7–13.
43. Weidner, H. & Crothers, D. M. (1977) *Nucleic Acids Res.* **4,** 3401–3414.
44. Walstrum, S. A. & Uhlenbeck, O. C. (1990) *Biochemistry* **29,** 10573–10576.
45. Herschlag, D. & Cech, T. R. (1990) *Biochemistry* **29,** 10159–10171.
46. Uhlenbeck, O. C. (1995) *RNA* **1,** 4–6.
47. Brockwell, D. J., Smith, D. A. & Radford, S. E. (2000) *Curr. Opin. Struct. Biol.* **10,** 16–25.