Biophysical Journal Volume 104 March 2013 1326–1337

Understanding the Role of Three-Dimensional Topology in Determining the Folding Intermediates of Group I Introns

Chunxia Chen,[†△] Somdeb Mitra,[‡△] Magdalena Jonikas,[§] Joshua Martin,[†¶] Michael Brenowitz,[‡] and Alain Laederach[†*]

[†]Department of Biology, University of North Carolina, Chapel Hill, North Carolina; [‡]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY; [§]Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Children's Hospital Boston, Boston, Massachusetts; and [¶]National Evolutionary Synthesis Center, Durham, North Carolina

ABSTRACT Many RNA molecules exert their biological function only after folding to unique three-dimensional structures. For long, noncoding RNA molecules, the complexity of finding the native topology can be a major impediment to correct folding to the biologically active structure. An RNA molecule may fold to a near-native structure but not be able to continue to the correct structure due to a topological barrier such as crossed strands or incorrectly stacked helices. Achieving the native conformation thus requires unfolding and refolding, resulting in a long-lived intermediate. We investigate the role of topology in the folding of two phylogenetically related catalytic group I introns, the *Twort* and *Azoarcus* group I ribozymes. The kinetic models describing the Mg^{2+}-mediated folding of these ribozymes were previously determined by time-resolved hydroxyl (•OH) radical footprinting. Two intermediates formed by parallel intermediates were resolved for each RNA. These data and analytical ultracentrifugation compaction analyses are used herein to constrain coarse-grained models of these folding intermediates as we investigate the role of nonnative topology in dictating the lifetime of the intermediates. Starting from an ensemble of unfolded conformations, we folded the RNA molecules by progressively adding native constraints to subdomains of the RNA defined by the •OH time-progress curves to simulate folding through the different kinetic pathways. We find that nonnative topologies (arrangement of helices) occur frequently in the folding simulations despite using only native constraints to drive the reaction, and that the initial conformation, rather than the folding pathway, is the major determinant of whether the RNA adopts nonnative topology during folding. From these analyses we conclude that biases in the initial conformation likely determine the relative flux through parallel RNA folding pathways.

INTRODUCTION

A characteristic common to many, if not most, large RNA folding reactions is that they proceed via multiple pathways encompassing highly populated intermediates (1–4). A contributing factor in the long life of folding intermediates is the propensity of RNA polymers to adopt conformations where nonnative tertiary contacts form transiently. The folding of these conformations stalls, thus slowing the overall folding to the biologically active conformation (5–8). It remains difficult to predict how RNAs misfold and to determine the molecular mechanism by which intermediate misfolding occurs (6,9–11).

RNA folding kinetics is highly dependent on the molecule being studied, the solution conditions of the folding reaction (e.g., cation type, valency, and concentration), the presence of exogenous molecules such as protein chaperones and correct secondary-structure annealing (3,12–17). To better understand the role of primary, secondary, and tertiary structure in RNA folding, we studied by time-resolved •OH radical footprinting the folding of three phylogenetically related group I ribozymes, *Azoarcus*, *Tetrahymena*, and *Twort* (1,5,18–22). These experimental studies were conducted under one solution condition (10 mM $MgCl_2$ and 100 mM KCl, at 25°C) so as to facilitate comparison of the contributions of contact strength and topology to the folding kinetics. Under experimental conditions similar to those used in this study, these RNA molecules commence their Mg^{2+}-mediated folding with their native secondary structure formed (6,23,24). This enables the analysis of tertiary contact formation to be analyzed without the complication of alternative secondary structures.

This study explores the structures of the intermediates identified in folding of *Azoarcus* and *Twort* ribozymes, which have the longest and shortest lifetimes, respectively. We use the crystal structures of the introns (25,26), the individual nucleotide-resolution solvent-accessible surface reported by •OH radical footprinting, and measures of global compaction to constrain simulations that model the three-dimensional structures of the folding intermediates. The models are generated using the coarse-grained and knowledge-based representation of RNA developed within the nucleic acid simulation tool (NAST). This new, to our knowledge, approach allows us to efficiently sample conformational space with unprecedented efficiency (27). Our melding of experimental constraints and simulation illuminates the structures of the folding intermediates and suggests that the initial conformational ensemble plays

Submitted December 8, 2012, and accepted for publication February 7, 2013.

△Chunxia Chen and Somdeb Mitra contributed equally to this work.

*Correspondence: alain@unc.edu

Editor: Kathleen Hall.

0006-3495/13/03/1326/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.02.007

a dominant role in dictating the course of an RNA molecule through its folding pathway. Our studies also investigate the effects of contact order formation on the resolution of RNA misfolds, which has been identified in recent studies as important (28–30).

MATERIALS AND METHODS

RNA preparation and analytical ultracentrifugation

The two ribozymes were in vitro transcribed using the Megascript T7 Kit (Ambion, Austin, TX), from two linear DNA templates encoding nucleotides 4–204 of the *Azoarcus* ribozyme (L-3; R. Russell, University of Texas Austin, 2010, personal communication) and 9–251 of the *Twort* ribozyme (25), each of which contained a T7 promoter sequence immediately upstream of the ribozyme sequence. The transcribed RNA molecules were purified either by 5% polyacrylamide gel electrophoresis or by the Megaclear kit (Ambion), ethanol precipitated and dissolved in CE buffer (10 mM potassium cacodylate and 0.1 mM EDTA, pH 7.3).

Analytical sedimentation-velocity analyses were conducted in a Beckman XL-I analytical ultracentrifuge with an An-60 Ti 4 hole rotor using the absorption optics. RNA stock solutions (10–50 μM) were diluted in CEK buffer (10 mM potassium cacodylate, 100 mM KCl, and 0.1 mM EDTA, pH 7.3) to the final RNA concentrations corresponding to A_{260} of 0.5–0.6 optical density. The RNAs were denatured by heating at 95°C for 2 min and cooled to 50°C for ~15 min. For the Mg^{2+} titrations, the appropriate amount of $MgCl_2$ was added to the RNA solution and incubation at 50°C was continued for 30 min. The samples were then equilibrated at the folding temperature (25°C) for 1 h and loaded along with the buffer reference into the double-sector centrifuge cells. The samples were centrifuged at 35,000 rpm, and 60 scans were collected for each sample. The scans were analyzed with the software DCDT+ (31). SEDNTERP was used to estimate the density and viscosity of the buffer. The partial specific volume ($\bar{v}$) and the hydration of the RNA molecule were assumed to be 0.53 cm^3/g and 0.59, respectively.

Kinetic modeling

The time-resolved hydroxyl radical footprinting data were modeled kinetically using Kinfold (13,14,32) as described in Mitra et al. (1). Briefly, *k*-means clustering was used to assign individual regions of protection (generally three to four consecutive nucleotides) to slow, medium, and fast (Fig. 1, *blue*, *red*, and *green*, respectively) progress curves. The cluster centroids, as determined by a Café Pick Me Up distance metric were used for kinetic model fitting using the previously described model enumeration approach (32). The best-fitting kinetic model, as determined by the distance metric to the cluster centroid data are reported in Fig. 1.

Simulation details

We used the coarse-grained knowledge-based RNA structure-modeling tool NAST (http://www.simtk.org/home/nast) as a platform for the proposed study. Each nucleotide is represented as a point centered at the C3′ atom. This representation significantly simplifies the computational complexity by reducing the number of possible interactions to be considered. Four types of interaction were considered in the energy function (27): 1), bonded

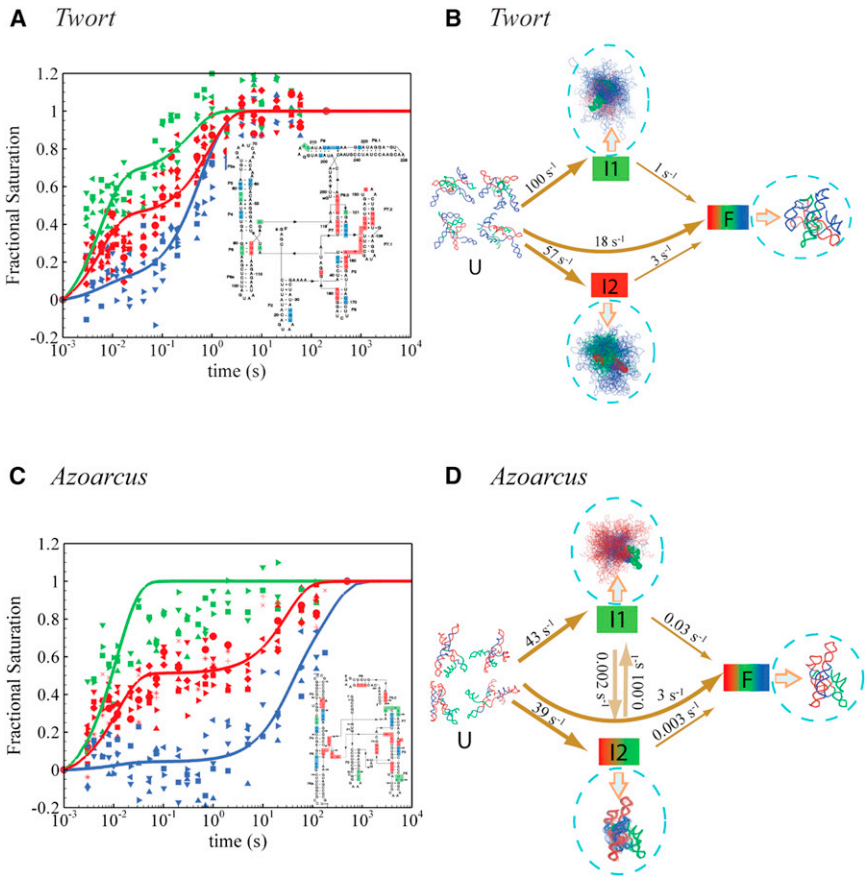

FIGURE 1 •OH radical data and resulting kinetic folding models for the *Twort* (*A* and *B*) and *Azoarcus* (*C* and *D*) group I introns studied in this manuscript. (*A* and *C*) Clustering of the time-resolved •OH footprinting data collected upon addition of 10 mM $MgCl_2$ at 37°C for the *Twort* and *Azoarcus* group I introns, respectively. The KinFold kinetic model fits to the data are plotted as solid lines through the clusters, and the colored insets illustrate the association of the domains in the molecule with the time-resolved data. (*B* and *D*) Kinetic models, determined by fitting the KinFold data in *A* and *C* (13,14,32), describing the folding reactions of the *Twort* and *Azoarcus* group I introns. Three-dimensional models for the unfolded (U) and intermediate (I) structures determined in this manuscript using the nucleic acid simulation tool (NAST) decorate the kinetic models. The process and analysis of the results of ensemble modeling of these RNA folding intermediates are described in the text.

interactions (distances, angles, and dihedrals between C3′ atoms of two, three, and four sequential nucleotides, respectively), which are derived from the geometries in RNA crystal structures; 2), repulsive nonbonded interactions; 3), ideal helical geometry constraints; and 4), long-range interaction between nucleotides participating in tertiary contacts (only used when folding RNAs). All the simulations were performed with a time step of 0.005 ps and at a temperature of 300 K. We randomly selected 100 misfolded structures for both *Azoarcus* and *Twort* introns. We performed 30,000 steps of NAST molecular dynamics to equilibrate these systems with secondary structures by releasing the tertiary interactions. Then 20,000 steps of molecular dynamics were performed to fold these equilibrated structures directly with tertiary interactions turned on. Global distance test, total score (GDT_TS), and root-mean-squared distance (RMSD) were calculated for the final structures to determine whether this process was able to resolve the misfold.

Computational generation of the unfolded ensemble

Our method for generating unfolded conformations constrains only the secondary structure to form native basepairs and includes no tertiary-structure contraints. The procedure is illustrated in Fig. 2 and involves first selecting a random residue in the primary sequence as the starting point. Two additional residues are iteratively added on each end of the strand at random positions 5.78 Å away from the first residue. We perform 1000 steps of NAST molecular dynamics to equilibrate the system in all cases. This process is repeated until all the residues have been added to the system. We repeated this procedure to generate 1000 starting conformations for both *Twort* and *Azoarcus* introns. In general, the mean of the radius of gyration (R_g) of these structures matches the experimentally determined radius of hydration (R_h). For the *Twort* intron, we did use a weak electrostatic repulsive force to increase our sampling of more extended structures to obtain an ensemble that precisely matches the experimental data.

Structural analysis of the unfolded ensemble

For both *Twort* and *Azoarcus*, we calculated the R_g by computing the mean pairwise distance between all atoms and averaging over the ensemble. Structural diversity was evaluated using the pairwise correlation values of the GDT_TS and RMSD for the ensemble. GDT_TS and RMSD are two scoring methods used to evaluate the similarity between two structures. GDT_TS measures the fraction of residues that are within distances of 1, 2, 4, and 8 Å from the residue positions in the target structure. RMSD calculates the residue-position distance of a structure from the target structure. A higher value of GDT_TS (in the range 0–1) and lower values of RMSD suggest well-matched structures, as previously reported (27). A cutoff of 5 Å for RMSD values was used to determine whether a final structure (F) successfully folded. This is based on an analysis of the *Azoarcus* I2 intermediate RMSD, where all intermediates leading to misfolded RNA had an RMSD greater than 5 Å (Fig. S5 *D* in the Supporting Material).

Folding simulations and intermediate modeling

The tertiary interactions between nucleotides were modeled using a spring potential, which is represented by a harmonic function. For each starting unfolded RNA structure, the folding was carried along different pathways:

1. Direct folding, where the tertiary interactions are turned on for all three domains simultaneously.
2. Folding through an intermediate structure, where the tertiary interactions are turned on successively to fold through different intermediates.

As shown in the kinetic models determined from experiments (Fig. 1, *C* and *D*), the *Twort* or *Azoarcus* introns each have two intermediates. To fold through I1, the green domain of the unfolded structure was constrained to native, and then all the three domains of the intermediate structure I1 were constrained to native to get the final folded structure. Likewise, to fold through I2, the red domain (for *Twort*, red-green domain for *Azoarcus*) was constrained to native and then all the three domains were constrained to native to get the final folded structure. Therefore, for both *Twort* and *Azoarcus*, we designed three folding pathways: fold through I1, fold

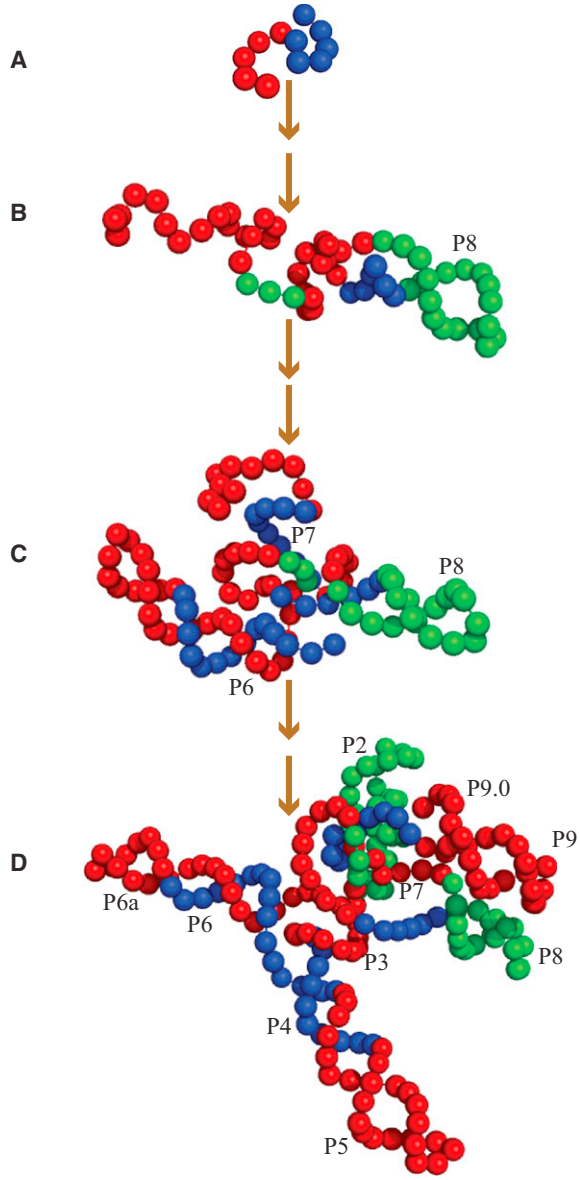

FIGURE 2 Flow chart illustrating the procedure for generating one (of 1000) unfolded conformation of the *Azoarcus* intron. RNA chain generation starts from one randomly selected nucleotide (in this case C131) and proceeds through a series of successive steps in which we add pairs of nucleotides to the 5′ and 3′ ends of the chain. Secondary structure is imposed using constraints previously described in Jonikas et al. (27). (*A*) Nucleotides G126–A136. (*B*) Nucleotides G102–G160, with P8 now fully formed. (*C*) Nucleotides G80–A182, with P6, P6a, P7, and P8 fully formed. (*D*) Final unfolded conformation with all the residues (total 195) added and the secondary structure fully formed. In this case, no native tertiary contacts are formed. We use an ensemble of unfolded conformations like these as a starting point for our folding simulations.

through I2, and fold directly. For each starting RNA structure, 20,000 steps of molecular dynamics simulations were performed to fold the structure through all three pathways, and one final folded structure is obtained for each folding pathway.

The folding simulations are driven by the application of native pairwise constraints on all coarse atoms of the molecule. The molecule is considered to fold successfully if its final conformation has an RMSD of 5 Å or less and a GDT_TS >0.6 at the conclusion of the simulation, typically 20,000 steps. The native constraints applied to facilitate folding result in the native state in only 44% of the simulations for *Twort* (Fig. S2 *A*) and 59% of the simulations for *Azoarcus* (Fig. S2 *B*). We performed further structural analysis on the misfolded intermediates to understand the specific interactions that prevented them from folding to native.

Intermediate structural analysis

As discussed previously, the intermediate structures are obtained through folding specific domains sequentially. To explore the conformational differences between intermediates that lead to successful or misfolded final structures in the folding pathway, for both *Twort* and *Azoarcus* introns, we identified these two types of intermediates by examining the GDT_TS and RMSD scores of the final folded structure relative to the crystal structure (>0.6 GDT_TS and <5 Å RMSD, folded, otherwise misfolded). We then overlaid these intermediates with the crystal structure by aligning the constrained regions of the RNA and calculated the average residue distance between intermediates and crystal structures. We then calculated the difference between the distances for folded and misfolded structures and performed a *z*-score analysis for every residue by normalizing the difference over variance for folded and misfolded structures.

RESULTS

Hydroxyl radical time-progress curves define kinetic folding models

Quantitative time-resolved hydroxyl radical (•OH) footprinting was used previously to experimentally resolve the kinetic models that describe the Mg^{2+}-mediated folding of the *Twort* and *Azoarcus* group I ribozyme folding reactions and identify the tertiary contacts present in the folding intermediates (Fig. 1) (1). This study extends that work by exploring the structures of the intermediate species resolved in the kinetic models (25,26). The folding experiments are briefly summarized here; details are provided in the previously published work (1). Time-progress curves were acquired in a quenched-flow mixer by initiating ribozyme folding with the addition of $MgCl_2$ to a concentration of 10 mM in a solution of 10 mM potassium cacodylate, 0.1 mM EDTA, and 100 mM KCl, pH 7.3, at 37°C. Folding progress was measured at time points ranging from a few milliseconds to several minutes. Both ribozymes are fully active under this experimental condition (1). The presence of 100 mM K^+ engenders a relaxed initial-state ensemble in which the repulsive Coulombic forces between the negatively charged backbones of the RNA helices are screened, resulting in greater conformational flexibility (9,33,34). As reviewed in the Discussion, the presence of a significant concentration of monovalent cations profoundly influences these in vitro folding reactions.

The time-progress curves report the changes in solvent accessibility (referred to as protections) over the course of the folding reaction for individuals or groups of nucleotides in the RNA. The majority of the protections that were analyzed report the formation of discrete tertiary contacts identified in the structures of the native RNA. The resolved time-progress curves are clustered to simplify the kinetic modeling and correlate the folding rates to regions of the RNA molecules in a model-independent fashion (32). Three clusters of time-progress curves, designated fast (*green*), medium (*red*), and slow (*blue*), were resolved for both the *Azoarcus* and *Twort* ribozymes (Fig. 1, *A* and *C*). Kinetic modeling of the clusters using KinFold (13,14,24,32) yielded a common two-intermediate kinetic-model configuration, albeit with distinct kinetic rates for each ribozyme (Fig. 1, *B* and *D*).

Our comparative analysis of the structures of the intermediates of the *Azoarcus* and *Twort* ribozymes is prompted by three aspects of their folding mechanisms. First, the lifetimes of the *Twort* intermediates are short relative to those of *Azoarcus* intermediates. Second, the two ribozymes show inverted orders of formation of peripheral and core tertiary contacts. For *Azoarcus*, the peripheral and internal tertiary contacts are associated with the medium (*red*) and slow (*blue*) clusters, respectively (Fig. 1, *A* and *C*, *insets*). In *Twort*, the peripheral and internal tertiary contacts are associated with the slow (*blue*) and medium (*red*) clusters, respectively (1). Third, affiliations of the progress-curve clusters with the folding intermediates were observed to be different for the two ribozymes (Fig. 1, *B* and *D*) (1). For *Twort*, the cluster assignments are orthogonal; only the green cluster is protected in I1 and only the red cluster is protected in I2 (Fig. 1 *B*). For *Azoarcus*, the green cluster is again associated with the I1 intermediate, whereas I2 is associated with *both* the green and red clusters (Fig. 1 *D*). Our goal is to provide structural correlates to these kinetic characteristics.

Sampling the unfolded ensemble

The focus of this study is the structural nature of the folding intermediates identified in the structural-kinetic models (Fig. 1). Our approach is to use the local changes in solvent accessibility reported by the •OH protections as constraints for coarse-grained modeling (27). In the first step of our analysis, we developed a strategy to generate diverse RNA conformations that model the unfolded ensemble that is the starting point for folding simulations. Fig. 2 illustrates this strategy for the *Azoarcus* ribozyme; the same strategy was used for the *Twort* RNA. To construct the initial ensemble, we randomly choose a nucleotide in the RNA and successively add pairs of nucleotides, one to each end of the chain. In Fig. 2 *A*, we have added five pairs of nucleotides (G126–A136) around nucleotide C131. As more pairs are added we constrain the secondary structure of the helical

regions to that of the native molecule. Fig. 2 *B* shows the formation of P8 after the addition of 20 bp. Minimization and equilibration steps are iteratively performed at each step to relax the structure. This process is repeated until an unfolded conformation is generated (Fig. 2, *C* and *D*).

Our strategy for generating the unfolded ensemble makes several assumptions. First, we assume that the RNA in the unfolded ensemble has native secondary structure before the formation of tertiary contacts (27); this assumption has been experimentally confirmed by chemical probing (6,23,24). We also constrain the ensemble to a mean radius of gyration (R_g), as measured by the average pairwise distance between all beads in our coarse-grained RNA model. R_g is experimentally constrained from analytical sedimentation-velocity analysis (31, 35–37) as measured at the initial condition of the kinetics folding experiments by the measured radius of hydration (R_h = 44.8 and 39.7 Å for the *Twort* and *Azoarcus* ribozymes, respectively; Table S1).

The histogram of R_g values for the 1000 *Twort* and *Azoarcus* ribozyme conformations that were generated reveals that the ensemble of unfolded RNA conformations agree with experimentally measured R_h values (Fig. 3, *A* and *B*, and Table S1). Nine representative unfolded conformations of different sizes (compact, mean, and extended) are shown in Fig. 3, *C* and *D*, to assist in the visualization of the unfolded ensemble. We measured a mean interconformation root-mean-square distance (RMSD) of >40 Å and an average global distance test (GDT_TS) below 0.1 for both the *Twort* and *Azoarcus* unfolded conformational ensembles (Fig. S1). A majority of the structures are different by at least 20 Å RMSD, demonstrating that our conformational sampling efficiently generates a diverse population of conformations (Fig. S1, *E* and *F*).

As an additional test of whether highly divergent structures are present in our ensemble, we generated an equivalent set of random-coil structures for the *Azoarcus* group I intron, i.e., the secondary structure was not constrained to native; this protocol yielded highly diverse structures (Fig. S7). We observe that when no constraints are used, the most divergent structures in terms of RMSD also have the largest R_g. RMSD and R_g values are highly correlated in this random-coil ensemble (Fig. S7 *A*). Our unfolded ensemble has structural diversity comparable to that of this random coil (Fig. S7 *B*), suggesting that although the ensemble of secondary-structure-constrained unfolded starting conformations is less diverse than a purely random-coil simulation, the initial ensemble is nonetheless composed of highly divergent structures.

Simulating folding along the kinetic pathways to model the intermediate structures

The NAST knowledge-based potential does not itself model specific tertiary contacts (27). Therefore, to drive the folding

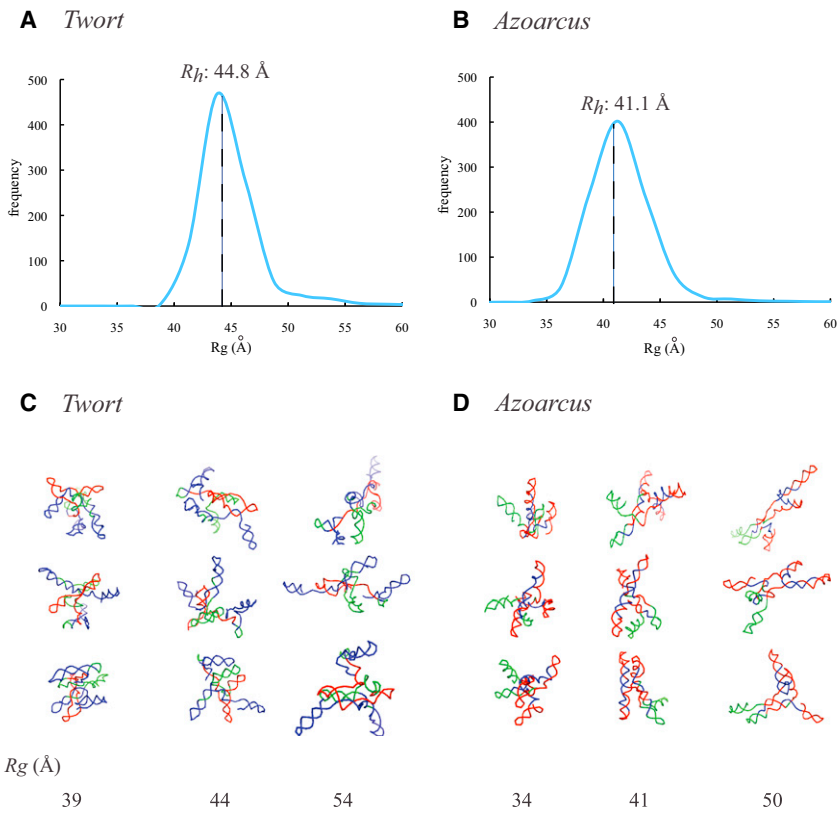

FIGURE 3 (*A* and *B*) R_g distribution of the unfolded conformation for the *Twort* and *Azoarcus* group I introns for 1000 conformations generated using NAST, a coarse-grained knowledge-based RNA structural modeling software (27). To generate these conformations, we applied constraints that impose a correct secondary structure to the model but do not otherwise define it. Experimental data (from analytical ultracentrifugation measurements) are also included in both figures for comparison. (*C* and *D*) Representative conformations illustrated for three different sizes of the *Twort* and *Azoarcus* group I introns.

reaction, we applied native constraints to the individual domains, identified by the •OH radical time-progress curve clustering (Fig. 1, *red, green, and blue*). We implemented these constraints as a network of weak harmonic springs between all the nucleotides in a given domain to drive the simulation toward nativelike conformations. These springs are 20-fold weaker than those used to constrain the secondary structure to A-form helix conformation and thus broadly mimic the relative strengths of these two classes of interactions.

Fig. 4 *A* illustrates this process schematically for one of the 1000 unfolded conformations of the *Twort* ribozyme. For example, to model folding through the I2 intermediate, we first applied constraints to the red domain, equilibrated the RNA with molecular dynamics, and then applied additional native constraints to the blue and green domains.

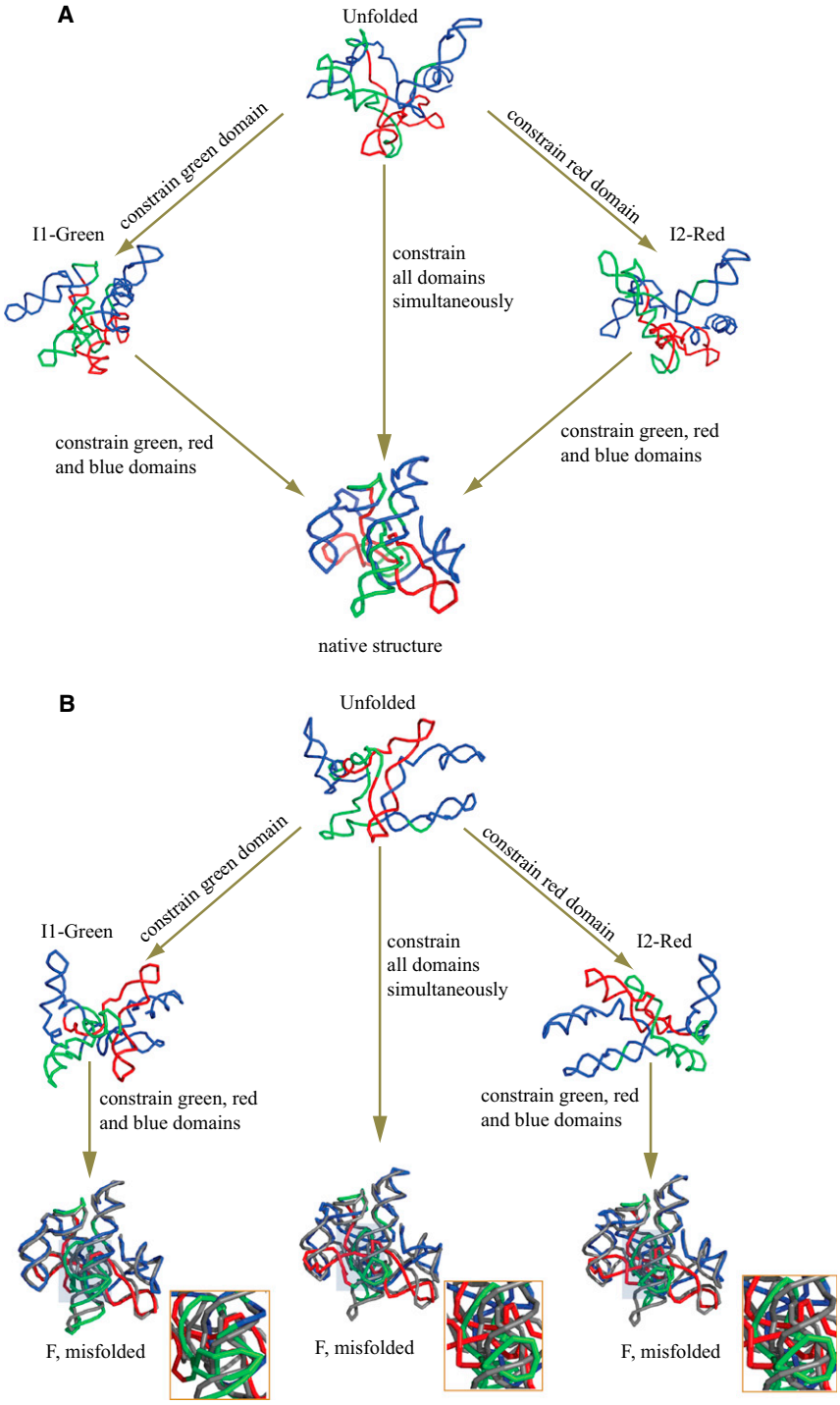

FIGURE 4 Examples of folding simulations starting from unfolded *Twort* that probe the intron's three major folding pathways. The intron is folded by successively applying native constraints to each domain (Fig. 1, *red*, *green*, and *blue*) and equilibrating the structures using coarse-grained molecular dynamics. (*A*) Folding through all three pathways leads to the native structure. (*B*) Folding through all three pathways leads to misfolded structures. The native structure is represented in gray to better visualize the misfolded state.

This procedure results in a correctly folded RNA (RMSD <5 Å and GDT_TS >0.6). Applying the same procedure to an alternative initial conformation of the same RNA (Fig. 4 *B*) yields a misfolded RNA (RMSD >5 Å and GDT_TS <0.6). The three pathways tested for the initial conformation illustrated in Fig. 4 *A* yield a correctly folded RNA. In contrast, the three pathways tested for the initial conformation illustrated in Fig. 4 *B* all lead to misfolded RNA.

We performed 3000 such folding simulations for each ribozyme (three pathways for each of the 1000 unfolded conformations) and determined the fraction of pathways that lead to correctly versus misfolded final states (Fig. S2). A correctly folded structure is defined as having an RMSD of <5 Å and a GDT_TS of >0.6 relative to the native conformation. A majority of initial conformations (83% for *Twort* and 73% for *Azoarcus*) fold or misfold independently of the pathway chosen (as illustrated in Fig. 4). These results reveal that the initial conformation significantly predetermines folding outcome. For a few unfolded starting conformations (17% and 27% for *Twort* and *Azoarcus*, respectively (Fig. S2)), the pathway chosen does influence folding outcome. An example of such a conformation is illustrated in Fig. S6, where folding through I1 yields a misfolded conformation, whereas folding through I2 or directly results in a properly folded structure.

We further characterized the 1000 unfolded conformations to better understand the structural features in the unfolded state that predispose the RNA to misfold. The size distributions of the unfolded conformations that either fold or misfold are the same within error, showing that whether the unfolded molecule is extended or compact does not affect folding outcome (Fig. S3, *A* and *B*). What differs between conformations that fold and misfold is their RMSDs relative to the crystal structure of the ribozyme (Fig. S3, *C* and *D*). Thus, size alone does not matter; it is the similarity of an unfolded conformation to the native structure that predisposes it to fold correctly.

Intermediate structures and their effect on topology

We explored the properties of the ensembles of the folding intermediates to understand the structural differences that lead to folded or misfolded RNA. As was the case for the unfolded ensembles, there is no difference in global size distribution between intermediates that correctly fold and those that misfolded. However, the RMSDs of correctly folding intermediates are smaller than RMSDs of misfolding intermediates (Figs. S4 and S5); formation of an intermediate structure closer to the native structure (lower RMSD) clearly increases the probability of successful folding. We also performed refolding simulations of misfolded intermediates to determine if the misfold could be easily resolved by allowing the conformation to equilibrate before reapplying the native folding potential. Interestingly, only 15 of 100 misfolded *Azoarcus* and 17 of 100 misfolded *Twort* intermediate conformations resolved to native folding. This result shows that the effect of the initial conformation is not easily decoupled from the folding pathway.

The conformational diversity of the RNA-folding-intermediate structural models is illustrated in the overlay of 50 representative structures of the *Twort* and *Azoarcus* I1 and I2 intermediates (Fig. 5). Except for the *Azoarcus* I2 intermediate, there is no difference in the structures of the ensembles that yield folded and misfolded RNA (Fig. 5, *F* and *H*). The conformations that result in misfolded *Azoarcus* I2 intermediate have a higher mean RMSD than those that fold when overlaid to the native crystal structure (Fig. 5, *F* and *H*). This result suggests that the *Azoarcus* ribozyme must adopt a specific conformation to allow correct formation of the catalytic core when folding through I2. The *Azoarcus* I2 intermediate is unique, because it is highly structured; only a specific conformation of Azoarcus I2 always misfolds. For the three other intermediates a variety of conformations can lead to misfolding. Therefore, the adoption of alternative three-dimensional topologies during the folding reaction is a common feature of all RNA.

Helix orientation as an indicator of alternative topology

To determine whether one or more helices are differently oriented in intermediates that either fold or misfold, we computed the average RMSD of individual helices within the intermediate models relative to the native (crystal) structure. We compared the differences in the means using a z-score (Table 1). The *z*-score comparison of the individual helices reveals two types of intermediate. The first type (*Twort* I1 and I2 and *Azoarcus* I1) has low *z*-scores (<0.6) for all of the helices. In the second type, *Azoarcus* I2, the majority of *z*-scores are >0.6. For P6, P6a, and P7 in *Azoarcus* I2, the *z*-scores are >1.0, indicating a significant difference in the orientation of the helices for intermediates that lead to folded and misfolded RNA. It is noteworthy that the *Azoarcus* I2 intermediate has the longest lifetime under our folding conditions (1). These results suggest that of the intermediates, only *Azoarcus* I2 adopts a systematically nonnative three-dimensional topology due to the incorrect orientation of the P6, P6a, and P7 helices.

DISCUSSION

Determining the number and nature of folding intermediates is essential to understanding the mechanism of a folding reaction that is more complex than two-state. Since the folding of large RNA molecules more often than not follows multiple pathways through multiple intermediates, knowledge of the structures of the intermediates is a prerequisite

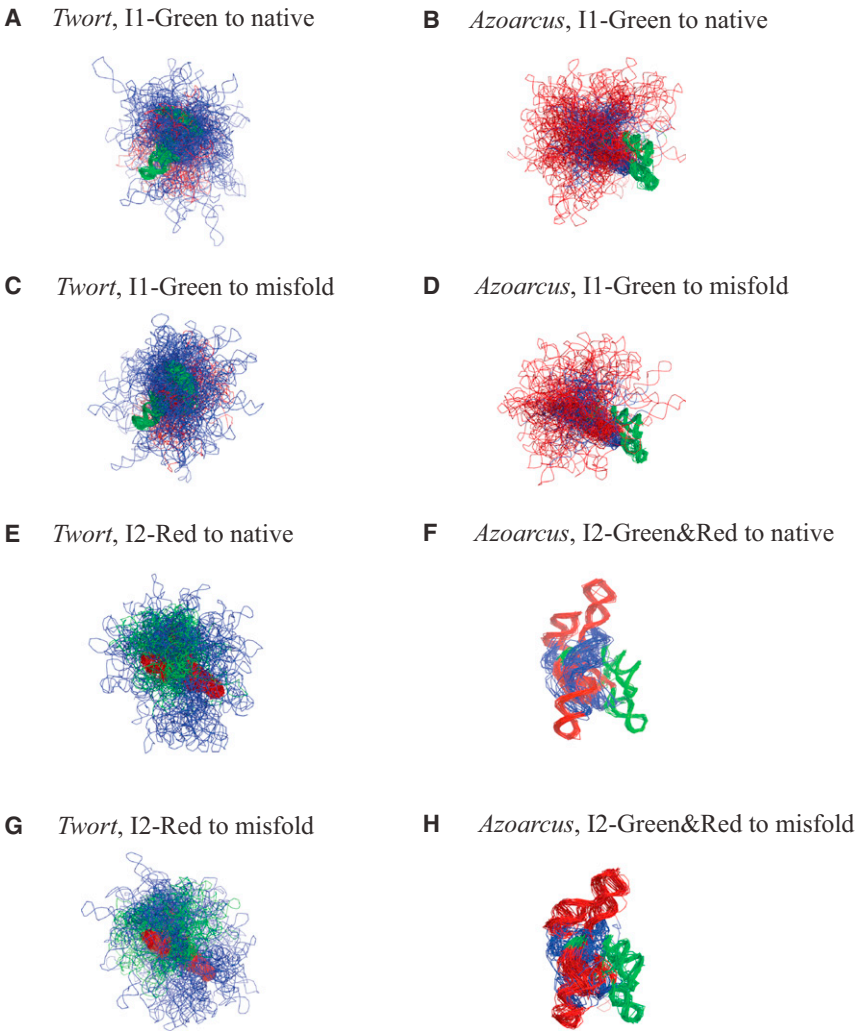

FIGURE 5 Overlay of 50 representative structural models for intermediate conformations of the *Twort* (*left column*) and *Azoarcus* (*right column*) group I introns that lead to either folded or misfolded structures when all native constraints are applied. Coloring of the domains is identical to the scheme in Fig. 1. (*A*) *Twort*, I1-green leading to native structure. (*B*) *Azoarcus*, I1-green leading to native structure. (*C*) *Twort*, I1-green leading to misfolded structure. (*D*) *Azoarcus*, I1-green leading to misfolded structure. (*E*) *Twort*, I2-red leading to native structure. (*F*) *Azoarcus*, I1-green and red leading to native structure. (*G*) *Twort*, I2-red leading to misfolded structure. (*H*) *Azoarcus*, I1-green and red leading to misfolded structure.

for completely describing an RNA folding mechanism. Structural-kinetic modeling of the time-progress curves obtained from time-resolved •OH footprinting analysis provides experimental insight into the number of folding intermediates and their interconnectivity, lifetimes, and structures (1). We leveraged the local changes in the solvent-accessible surface of the phosphodiester backbone captured in the time-progress curves to model the intermediates of two RNA folding reactions. Our study of folding intermediate structures reveals the primacy of the initial conformation and the secondary nature of intermediate topology in determining the overall time course of a folding reaction.

The intermediate structures resolved from our study provide significant insight into mechanisms of RNA misfolding (Fig. 5). We conclude from our statistical analysis of relative helix orientations that RNAs readily adopt alternative conformations and that alternative three-dimensional topology is a common event in the folding reactions of the *Twort* and *Azoarcus* ribozymes; helices can and do become tangled given certain initial conformations (Table 1). In the cases of *Twort* I1 and I2 and *Azoarcus* I1, this process is not necessarily helix-specific (the *z*-scores in Table 1 are not significant). Therefore, RNA molecules with stable secondary structure will have a tendency to adopt nonnative conformations regardless of the detailed arrangement of their secondary structure. It is only in specific cases, usually when a strong tertiary contact is formed, that specific intermediate structures misfold in a well-defined manner. This is the case for *Azoarcus* I2, where the P6, P6a, and P7 helices are incorrectly positioned in the intermediate relative to the native structure (Fig. 5 *H*).

It has been theorized that similar misfolding occurs in the *Tetrahymena* ribozyme (6). The folding energy landscape for RNA (and in some cases proteins) is often described as rugged, which suggests that the initial conformation will play an important role in determining folding outcome (37–43). In addition, our structural modeling data support the hypothesis that alternative three-dimensional topological organization leads to long-lived misfolded intermediates in multiple group I introns (23). Our simulations are therefore consistent with a rugged energy landscape for

TABLE 1 *z*-Score and residue-distance difference between folded and misfolded intermediates for *Twort* and *Azoarcus*

	Twort		*Azoarcus*	
Domain name	I_1-Green	I_2-Red	I_1-Green	I_2-Red/Green
P_2	0.03 (0.01 Å)	0.55 (−15.11 Å)	0.06 (−0.09 Å)	0.63 (−3.46 Å)
P_3	0.10 (−0.44 Å)	0.21 (−3.33 Å)	0.09 (−0.88 Å)	0.97 (−7.90 Å)
P_4	0.11 (−0.71 Å)	0.36 (−5.93 Å)	0.09 (−4.74 Å)	0.78 (−7.10 Å)
P_5	0.12 (2.24 Å)	0.20 (−5.50 Å)	0.11 (3.93 Å)	0.99 (−4.13 Å)
P_{5a}	0.16 (4.31 Å)	0.32 (−11.28 Å)		
P_6	0.22 (−3.41 Å)	0.07 (−0.20 Å)	0.44 (−6.95 Å)	1.12 (−9.22 Å)
P_{6a}	0.23 (−5.67 Å)	0.21 (−1.32 Å)	0.45 (−10.68 Å)	1.30 (−10.79 Å)
P_7	0.07 (0.37 Å)	0.29 (−1.87 Å)	0.03 (−0.46 Å)	1.27 (−7.17 Å)
$P_{7.1}$	0.04 (−0.54 Å)	0.11 (0.12 Å)		
$P_{7.2}$	0.13 (−0.23 Å)	0.14 (−1.89 Å)		
P_8	0.09 (−0.45 Å)	0.36 (−9.18 Å)	0.08 (−0.26 Å)	0.55 (−3.53 Å)
$P_{9.0}$	0.12 (−1.25 Å)	0.37 (−3.96 Å)	0.07 (−1.26 Å)	0.85 (−7.94 Å)
P_9	0.17 (2.54 Å)	0.38 (−7.40 Å)	0.02 (0.48 Å)	0.66 (−2.35 Å)
$P_{9.1}$	0.18 (4.23 Å)	0.54 (−14.23 Å)		

RNA folding, but they provide more details regarding the specific molecular interactions that result in misfolding. Furthermore, a conclusion drawn from the experimental time-resolved •OH folding experiments is that the order in which the peripheral and core tertiary contacts form is inverted for the folding of the *Twort* and *Azoarcus* ribozymes under these experimental questions (1). Although this observation is extensively discussed elsewhere (1), it should be noted that the most structurally uniform intermediate, that of *Azoarcus* I2, occurs in a molecule where the peripheral contacts form before the core in a way that is similar to that observed in the analogously folding *Tetrahymena* group I intron (6,13,14,35,44,45).

An assumption made in our structural modeling is that each •OH radical protection reports the formation of a native tertiary contact. However, we observe in our folding simulations misfolded structures despite the application of weak native constraints that reflect the experimental constraints (Fig. 4 *B*, where nonnative contacts are clearly present). We purposely chose to impose weak native constraints to allow the formation of nonnative contacts. Our results illustrate the extent to which RNA is susceptible to misfolding and that imposition of weak native constraints in a folding simulation does not exclude formation of nonnative contacts. This approach is an improvement over our previous interpretation of the structural information inherent to •OH time-progress curves, which simply mapped the •OH protections to the native structure (13).

Flux analyses of the experimentally derived kinetic models (Fig. 1, *B* and *D*) predict that only a small fraction of the molecules follow the direct U→F pathway, 10% and 3% for *Twort* and *Azoarcus*, respectively (1). The flux calculated from our simulations that passes through the direct pathway shows how the simulations systematically overestimate this pathway (Fig. S2, *A* and *B*; 44% and 59% for *Twort* and *Azoarcus*, respectively). This discrepancy is likely the result of our use of purely native constraints in folding simulations. By definition, the imposition of native constraints will favor folding to the native structure. However, nearly half of the starting conformations for both RNAs lead to at least one misfolded intermediate despite this assumption. This result is unsurprising, since it is well established that nonnative interactions form during RNA folding reactions (6,11,41,47). Although we could have used nonnative constraints in our simulations to increase the flux to misfolded intermediates, we are specifically interested in the RNA polymer's intrinsic propensity to misfold. Our use of the simpler, and thus more tractable, folding model yields the sought insight into misfolding, with the collateral consequence of not completely capturing the experimentally observed folding behavior. Having established the intrinsic propensity of these RNA molecules to misfold, it would be worthwhile in future studies to more fully model folding by imposing both native and nonnative constraints.

Our results are in agreement with those of single-molecule studies of the P4–P6 domain of the *Tetrahymena* ribozyme (41) that show the importance of the initial conformation in determining RNA folding dynamics. These experiments revealed that identical single molecules have very different and discrete kinetic behaviors. The discrete classes of molecules rarely interchanged, indicating that their behavior is dictated by the initial conformation before the addition of $MgCl_2$. This observation is consistent with our simulation results, as is the observation that neither the pathway taken nor the relaxation of tertiary constraints easily resolve three of the four misfolded intermediates that we studied. In addition, we observe structural correlations within our sampling of the unfolded ensemble, suggesting that secondary-structure constraints also impose at least a minimal level of tertiary structure (Fig. S1).

Published *Azoarcus* folding studies carried out from a low-salt initial condition (48,49) yield folding rates much faster than those measured starting at the higher salt concentration of 100 mM KCl used in our experiment (1). That the initial folding condition significantly impacts the

kinetics native structure formation was also observed for the *Tetrahymena* group I ribozyme, although in this case, Mg^{2+}-mediated folding is slower when initiated from the low-salt initial condition (50). Our conclusion that the majority of conformations predispose the RNA to either fold or misfold, regardless of which pathway is chosen, is consistent with the importance of the initial conditions in determining folding rate and flux through the different pathways (14). By selecting initial conditions that favor folding over misfolding, RNAs can be predisposed to fold faster. For *Azoarcus* at the low salt initial condition, the highly extended conformations are presumably less likely to allow early formation of the strong tetraloop-tetraloop receptor interactions that favor misfolds in our I2 structural model (Fig. 5 *H*). Although we do not model ionic conditions in our simulations, which significantly affects folding flux, as discussed above (12,51–54), our simulations suggest that ionic conditions likely influence the relative abundance of specific conformations in the initial ensemble (34,49,55).

Time-resolved small-angle x-ray scattering studies of group I intron folding clearly show a rapid collapse upon addition of a divalent cation that is analogous in scope to our addition of native constraints (54,56,57). To test this conjecture, we performed additional reequilibration simulations on the 83 *Azoarcus* conformations that did not resolve to native in the first step; this procedure further resolved seven misfolded structures. When we doubled the equilibration time, we were able to refold an additional nine structures. This result suggests that successive reequilibration steps would eventually resolve a majority of RNA molecules in the simulations. The fact that reequilibrating misfolded intermediates for a short time only resolves a moderate percentage of the conformations further illustrates the importance of the starting conformation in determining folding outcome.

The disagreement between the flux through simulated folding pathways reported in Fig. S2 and that determined experimentally through analysis of the kinetic parameters of the folding models (Fig. 1, *B* and *D*) is large. Factors other than three-dimensional topology may account for the discrepancy between the calculated and experimentally determined flux. One major assumption of our structural modeling is uniform secondary structure of the starting ensemble, which is supported by chemical probing data of both ribozymes in the absence of $MgCl_2$, although under slightly different monovalent conditions for the *Azoarcus* ribozyme (24,48,49). RNA molecules with nonnative secondary structures will likely fold slowly through populated intermediates given the significant barrier associated with changing secondary structure in an RNA (6,7,58). A recent study showing that the P3/P7 pseudoknot in the *Azoarcus* ribozyme readily misfolds highlights that secondary structure plays a central role in folding outcome (59). In addition, chemical probing experiments cannot entirely rule out the presence of a small fraction of nonnative secondary structures that would contribute to the discrepancy between our flux simulations and the experimental measurements (24). As such, the discrepancies we observe between experimental and simulated flux may also be due to small subpopulations of RNAs adopting alternative secondary structures. Although these types of differences could be incorporated in our models and would likely result in misfolded intermediates (as per the structural definition used in this manuscript), one limiting factor is that there is no experimental evidence for the precise basepairs of these alternative secondary structures. As such, including them in this type of study would require assumptions even beyond the ones we made here.

In conclusion, we have effectively leveraged in our study recent developments in both kinetic and structural modeling to address the role of topology in RNA folding. We interrogated the first step of two folding reactions and found that the initial conformation is the dominant determinant of whether RNA will misfold. Our conclusions are in agreement with those drawn from recent small angle x-ray scattering experiments (60). A folding intermediate will ultimately resolve a misfolded state by unfolding to a state that is predisposed to correct folding. However, this process takes time, which rationalizes the long life of misfolded RNA intermediates. Our work also reveals a very large difference in dynamic behavior for two otherwise structurally and phylogenetically similar RNA molecules. From an evolutionary perspective, RNA's dynamic diversity makes it an ideal molecule to effectively sample phenotypic space with limited genotypic variation. This characteristic is likely a major contributor to RNA's fulfillment of its diverse biological roles in the cell.

SUPPORTING MATERIAL

Seven figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00192-6.

The authors wish to thank Russ Altman and Randy Radmer for their assistance in the development of NAST.

This work was supported by the National Institutes of Health through an ARRA supplement to K99/R00-GM079953, and grants R21-MH087336, R01-HL111527, and R01-GM101237 to A.L., and RO1-GM085130 to M.B.

REFERENCES

1. Mitra, S., A. Laederach, …, M. Brenowitz. 2011. RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways. *RNA.* 17:1589–1603.
2. Woodson, S. A. 2008. RNA folding and ribosome assembly. *Curr. Opin. Chem. Biol.* 12:667–673.
3. Woodson, S. A. 2000. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* 57:796–808.
4. Pan, J., D. Thirumalai, and S. A. Woodson. 1997. Folding of RNA involves parallel pathways. *J. Mol. Biol.* 273:7–13.

5. Thirumalai, D., and C. Hyeon. 2005. RNA and protein folding: common themes and variations. *Biochemistry.* 44:4957–4970.
6. Russell, R., R. Das, …, D. Herschlag. 2006. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J. Mol. Biol.* 363:531–544.
7. Lescoute, A., and E. Westhof. 2006. Topology of three-way junctions in folded RNAs. *RNA.* 12:83–93.
8. Jewett, A. I., V. S. Pande, and K. W. Plaxco. 2003. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J. Mol. Biol.* 326:247–253.
9. Takamoto, K., R. Das, …, M. R. Chance. 2004. Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J. Mol. Biol.* 343: 1195–1206.
10. Lipfert, J., A. Y. L. Sim, …, S. Doniach. 2010. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA.* 16:708–719.
11. Forconi, M., J. A. Piccirilli, and D. Herschlag. 2007. Modulation of individual steps in group I intron catalysis by a peripheral metal ion. *RNA.* 13:1656–1667.
12. Silverman, S. K., M. L. Deras, …, T. R. Cech. 2000. Multiple folding pathways for the P4-P6 RNA domain. *Biochemistry.* 39:12465–12475.
13. Laederach, A., I. Shcherbakova, …, R. B. Altman. 2006. Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule. *J. Mol. Biol.* 358:1179–1190.
14. Laederach, A., I. Shcherbakova, …, M. Brenowitz. 2007. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc. Natl. Acad. Sci. USA.* 104:7045–7050.
15. Brenowitz, M., M. R. Chance, …, K. Takamoto. 2002. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting". *Curr. Opin. Struct. Biol.* 12:648–653.
16. Adilakshmi, T., P. Ramaswamy, and S. A. Woodson. 2005. Protein-independent folding pathway of the 16S rRNA 5′ domain. *J. Mol. Biol.* 351:508–519.
17. Adilakshmi, T., D. L. Bellur, and S. A. Woodson. 2008. Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature.* 455:1268–1272.
18. Talkington, M. W., G. Siuzdak, and J. R. Williamson. 2005. An assembly landscape for the 30S ribosomal subunit. *Nature.* 438: 628–632.
19. Thirumalai, D., and S. A. Woodson. 2000. Maximizing RNA folding rates: a balancing act. *RNA.* 6:790–794.
20. Treiber, D. K., M. S. Rook, …, J. R. Williamson. 1998. Kinetic intermediates trapped by native interactions in RNA folding. *Science.* 279:1943–1946.
21. Zarrinkar, P. P., and J. R. Williamson. 1994. Kinetic intermediates in RNA folding. *Science.* 265:918–924.
22. Zarrinkar, P. P., and J. R. Williamson. 1996. The kinetic folding pathway of the *Tetrahymena* ribozyme reveals possible similarities between RNA and protein folding. *Nat. Struct. Biol.* 3:432–438.
23. Russell, R., X. Zhuang, …, D. Herschlag. 2002. Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA.* 99: 155–160.
24. Quarrier, S., J. S. Martin, …, A. Laederach. 2010. Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA.* 16:1108–1117.
25. Golden, B. L., H. Kim, and E. Chase. 2005. Crystal structure of a phage *Twort* group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* 12:82–89.
26. Adams, P. L., M. R. Stahley, …, S. A. Strobel. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature.* 430:45–50.
27. Jonikas, M. A., R. J. Radmer, …, R. B. Altman. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA.* 15:189–199.
28. Quince, C., A. Lanzen, …, P. J. Turnbaugh. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics.* 12:38.
29. Sinan, S., X. Yuan, and R. Russell. 2011. The *Azoarcus* group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA chaperone proteins. *J. Biol. Chem.* 286:37304–37312.
30. Russell, R., I. Jarmoskaite, and A. M. Lambowitz. 2012. Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol.* 10: Epub ahead of print, September 20, 2012.
31. Philo, J. S. 2006. Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* 354:238–246.
32. Martin, J. S., K. Simmons, and A. Laederach. 2009. Exhaustive enumeration of kinetic model topologies for the analysis of time-resolved RNA folding. *Algorithms.* 2:200–214.
33. Russell, R., I. S. Millett, …, D. Herschlag. 2000. Small angle x-ray scattering reveals a compact intermediate in RNA folding. *Nat. Struct. Biol.* 7:367–370.
34. Rangan, P., B. Masquida, …, S. A. Woodson. 2003. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. USA.* 100:1574–1579.
35. Takamoto, K., Q. He, …, M. Brenowitz. 2002. Monovalent cations mediate formation of native tertiary structure of the *Tetrahymena thermophila* ribozyme. *Nat. Struct. Biol.* 9:928–933.
36. Philo, J. S. 1997. An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys. J.* 72: 435–444.
37. Bokinsky, G., and X. Zhuang. 2005. Single-molecule RNA folding. *Acc. Chem. Res.* 38:566–573.
38. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
39. Rook, M. S., D. K. Treiber, and J. R. Williamson. 1998. Fast folding mutants of the *Tetrahymena* group I ribozyme reveal a rugged folding energy landscape. *J. Mol. Biol.* 281:609–620.
40. Shcherbakova, I., S. Mitra, …, M. Brenowitz. 2008. Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr. Opin. Chem. Biol.* 12:655–666.
41. Solomatin, S. V., M. Greenfeld, …, D. Herschlag. 2010. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature.* 463:681–684.
42. Treiber, D. K., and J. R. Williamson. 2001. Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction. *J. Mol. Biol.* 305:11–21.
43. Treiber, D. K., and J. R. Williamson. 2001. Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* 11:309–314.
44. Woodson, S. A. 2002. Folding mechanisms of group I ribozymes: role of stability and contact order. *Biochem. Soc. Trans.* 30:1166–1169.
45. Vicens, Q., A. R. Gooding, …, T. R. Cech. 2007. Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA.* 13:536–548.
46. Reference deleted in proof.
47. Chu, V. B., and D. Herschlag. 2008. Unwinding RNA's secrets: advances in the biology, physics, and modeling of complex RNAs. *Curr. Opin. Struct. Biol.* 18:305–314.
48. Chauhan, S., and S. A. Woodson. 2008. Tertiary interactions determine the accuracy of RNA folding. *J. Am. Chem. Soc.* 130:1296–1303.
49. Chauhan, S., R. Behrouzi, …, S. A. Woodson. 2009. Structural rearrangements linked to global folding pathways of the *Azoarcus* group I ribozyme. *J. Mol. Biol.* 386:1167–1178.
50. Laederach, A., J. M. Chan, …, R. B. Altman. 2007. Coplanar and coaxial orientations of RNA bases and helices. *RNA.* 13:643–650.

51. Sorin, E. J., B. J. Nakatani, …, V. S. Pande. 2004. Does native state topology determine the RNA folding mechanism? *J. Mol. Biol.* 337:789–797.

52. Serebrov, V., R. J. Clarke, …, L. Kisselev. 2001. Mg^{2+}-induced tRNA folding. *Biochemistry.* 40:6688–6698.

53. Heilman-Miller, S. L., D. Thirumalai, and S. A. Woodson. 2001. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. *J. Mol. Biol.* 306:1157–1166.

54. Das, R., L. W. Kwok, …, D. Herschlag. 2003. The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the *Tetrahymena* ribozyme. *J. Mol. Biol.* 332:311–319.

55. Rangan, P., B. Masquida, …, S. A. Woodson. 2004. Architecture and folding mechanism of the *Azoarcus* Group I Pre-tRNA. *J. Mol. Biol.* 339:41–51.

56. Schlatterer, J. C., L. W. Kwok, …, L. Pollack. 2008. Hinge stiffness is a barrier to RNA folding. *J. Mol. Biol.* 379:859–870.

57. Russell, R., I. S. Millett, …, L. Pollack. 2002. Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci. USA.* 99:4266–4271.

58. Whitford, P. C., A. Schug, …, K. Y. Sanbonmatsu. 2009. Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function. *Biophys. J.* 96:L7–L9.

59. Zhang, L., P. Bao, …, Y. Zhang. 2009. Slow formation of a pseudoknot structure is rate limiting in the productive co-transcriptional folding of the self-splicing *Candida* intron. *RNA.* 15:1986–1992.

60. Roh, J. H., L. Guo, …, S. A. Woodson. 2010. Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle x-ray scattering. *J. Am. Chem. Soc.* 132:10148–10154.