



Thermodynamics and Kinetics of the Hairpin Ribozyme from Atomistic Folding/Unfolding Simulations

Citation

Nivón, Lucas G., and Eugene I. Shakhnovich. 2011. "Thermodynamics and Kinetics of the Hairpin Ribozyme from Atomistic Folding/Unfolding Simulations." *Journal of Molecular Biology* 411 (5) (September): 1128–1144. doi:10.1016/j.jmb.2011.06.042.

Published Version

10.1016/j.jmb.2011.06.042

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33464146>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

J Mol Biol. 2011 September 2; 411(5): 1128–1144. doi:10.1016/j.jmb.2011.06.042.

Thermodynamics and kinetics of the hairpin ribozyme from atomistic folding/unfolding simulations

Lucas G. Nivón^{*,§} and Eugene I. Shakhnovich^{§,♦}

^{*}Program in Biophysics, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[§]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

Abstract

We report a set of atomistic folding/unfolding simulations for the hairpin ribozyme using a monte carlo algorithm. The hairpin ribozyme folds in solution and catalyzes self-cleavage or ligation via a specific two-domain structure. The minimal active ribozyme has been studied extensively, showing stabilization of the active structure by cations and dynamic motion of the active structure. Here we introduce a simple model of tertiary structure formation that leads to a phase diagram for the RNA as a function of temperature and tertiary structure strength. We then employ this model to capture many folding/unfolding events and to examine the transition state ensemble (TSE) of the RNA during folding to its active “docked” conformation. The TSE is compact but with few tertiary interactions formed, in agreement with single-molecule dynamics experiments. To compare with experimental kinetic parameters we introduce a novel method to benchmark monte carlo kinetic parameters to docking/undocking rates collected over many single molecular trajectories. We find that topology alone, as encoded in a biased potential which discriminates between secondary and tertiary interactions, is sufficient to predict the thermodynamic behavior and kinetic folding pathway of the hairpin ribozyme. This method should be useful in predicting folding transition states for many natural or man-made RNA tertiary structures.

Introduction

Ribozymes perform a variety of catalytic functions in nature and the exact three-dimensional structure in the vicinity of the active site is responsible for the vast rate-enhancements affected by RNA. It is therefore of great interest to understand the exact sequence of events through which a newly synthesized RNA chain folds into the catalytically active native state before commencing cleavage, ligation, or one of many other catalytic roles. Efforts to understand RNA folding mechanism have focused largely on auto-catalytic introns, such as the Group I intron from *Tetrahymena*, and on smaller self-cleaving RNAs such as the hairpin, hammerhead, hepatitis delta, and Neurospora VS ribozymes.^{1,2} The hairpin ribozyme was discovered in the negative strand of tobacco ringspot virus roughly twenty years ago.^{3,4} It has at least two domains of secondary structure joined by tertiary interactions in the active form.^{5,6} In the minimal, two-domain form, this ribozyme provides an ideal model for the formation of tertiary structure in RNA folding because it has so few

© 2011 Elsevier Ltd. All rights reserved.

[♦]Corresponding Author: eugene@belok.harvard.edu.

^{*}Present Address: Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

interacting components. Experimental work has focused most closely on the “docking” of domains to form the tertiary structure of the active ribozyme, which is critical for catalysis.⁷ Single-molecule studies have measured the rates of docking and undocking for the minimal form in different cations, and even pinpointed the atoms involved in stabilization of the transition state for folding.^{8,9}

RNA folding has been simulated for a pair of cases at atomic detail, and more recently Dokholyan and co-workers used a centroid model to fold and predict the structures of a number of RNA tertiary structures.¹⁰ The RNA folding process has only been simulated at atomic detail for a pair of cases: The GCAA tetraloop RNA^{11, 12, 13, 14, 15} and the tRNA^{Phe}.¹⁶ The smaller tetraloop was studied using molecular dynamics with both implicit and explicit solvent,^{12,13,14, 15} but collecting many folding simulations for a larger system like the hairpin ribozyme or tRNA^{Phe} would be extremely difficult without a Go-type potential. Both the tRNA and tetraloop have been studied using Go and Go-like potentials, in which only native interactions are considered, such that atom-atom contacts which are present in the crystal structure are attractive in any conformation and contacts not present in the folded state are repulsive.¹⁷ A standard Go potential encodes the topology of a bio-molecular structure in the form of a potential that assumes that contacts observed in the native state are the only ones that are energetically favorable, while a Go-like or biasing potential also includes other energetic contributions that may or may not favor the native state. By using such a potential, one can ask whether the folding pathway is determined primarily by topology. A Go-type potential implicitly includes solvation (surface atoms do not form native interactions and therefore cannot form energetically favorable interactions with other atoms) and rejects many non-native conformations, and therefore greatly accelerates the folding simulation process. It has been shown that topology plays a key role in the folding of many proteins.^{18, 19, 20} Furthermore, a topological analysis similar to that used for proteins has shown that topology is a strong predictor of RNA folding rate for “slow folding” RNAs, including the two-way junction hairpin ribozyme studied here.²¹ The details of the application of Go and Go-like potentials to RNA are discussed more fully in our earlier work using a pure Go potential to fold a smaller RNA and in the work of Pande and co-workers.^{11, 16}

Atomistic Go simulations have been used successfully in protein folding to predict the structure of a transition state, for example finding the conserved core residues in the folding nucleus of protein G.^{22, 23} Go simulations have also predicted intermediates in the folding of three-state folders barnase, RNase H and CheY.²⁴ Simulations combining phi-value constraints with Go potentials have verified and refined models of the TSE and identified the most critical residues in the folding nucleus.^{25, 26, 27} In the case of nucleic acids, Go simulations predict the correct “zipping” pathway for a GCAA tetraloop hairpin RNA¹¹, and a combined Go-type/AMBER94 simulation of tRNA predicted intermediates along the folding path in qualitative agreement with experimental work by Sosnick.^{16, 28}

The minimal hairpin ribozyme (two-way junction, 2WJ) has fewer domains than tRNA, and the thermodynamics and kinetics of tertiary structure formation have been well characterized for both the minimal and full (four way junction, 4WJ) forms of the RNA. In addition, the transition state for tertiary structure formation has been analyzed in some detail using single-molecule techniques.⁹ Theoretical studies of the folding of the ribozyme have used the nonlinear Poisson-Boltzmann equation to estimate the magnesium uptake of various coarse-grained hypothesized transition state structures.⁹ With the advent of techniques for the simulation of larger RNAs, it is now of great interest to simulate the full folding trajectory of the hairpin ribozyme at atomistic resolution using a Go-like potential.

Cations play a vital role in the formation of RNA tertiary structure. Tertiary structure formation necessarily involves placing negatively charged backbone phosphates in close proximity with one another. Monovalent cations can mitigate the repulsion between negative charges, allowing folding to proceed, but divalent cations do so more efficiently and appear to be necessary for full formation of tertiary structure in many, but not all, cases.^{29,30} Early work by Crothers and co-workers on tRNA demonstrated the existence of three major phases as a function of magnesium and temperature.³¹ In this phase diagram the RNA shows native (folded) structure at high cation concentration and low temperature, coil at high temperature (completely unfolded), and “extended forms” at low temperature and cation concentration. More recently, similar results have been observed for the hairpin ribozyme using fluorescence techniques, demonstrating a dynamic phase (interpreted as docking/undocking of individual helices) at low magnesium, entirely unfolded structures at high temperature³², and a stably docked state at higher magnesium and low temperature.⁹

The interaction of magnesium with RNA has been modeled as a combination of specifically bound ions along with a diffuse cloud surrounding the RNA, both of which act to preferentially stabilize compact RNA structures at higher cation concentrations.³³ This framework allows a detailed description of the influence of Mg^{2+} on the stability of various RNAs in unfolded, intermediate and native states. Such models of the role of Mg^{2+} in RNA thermodynamics have not been extended to the study of RNA dynamics or folding, but rather have focused on predicting the stability and concentration of bound ions for static structures. In the simulation of the tRNA^{Phe} the effects of Mg^{2+} were modeled in a fashion very similar to the present work, separating contacts into secondary or tertiary categories.¹⁶ Secondary interactions are those within an RNA helix while tertiary ones are between helices. Here we extend that analysis by studying the influence of varying the ratio of tertiary/secondary interaction strengths within a Go-like model on RNA thermodynamics – *Can the effect of varying cation concentration on RNA thermodynamics can be captured in a simple go-like model?*

The hairpin ribozyme is known to undergo a structural transition between the docked and undocked states, and this transition is recapitulated in our simulations of the ribozyme. We use the atomistic nature of these simulations to build a simulated model of the transition state for folding (or docking) of the ribozyme from an intermediate with secondary structure formed to the native state. The structure of the transition state for docking has been studied in previous single-molecule work by targeted mutations, salt titration, and urea titration.⁹ The structure of the TSE inferred from these experimental studies agrees well with our present findings.

We analyze the structure of the TSE by conducting repeated folding simulations starting from the undocked conformation. By adjusting τ , one stabilizes or destabilizes tertiary structure. By determining the folding behavior as a function of τ , we learn about the relative importance of tertiary structure in the TS – an approach similar in spirit to the ϕ -value analysis in protein folding.³⁴ Then, by carrying out further simulations, we identify an atomically detailed model of the TSE using P_{fold} simulations³⁵ similar to those used for the analysis of the folding of the GCAA tetraloop RNA.¹¹

The modified Go-type model in this work allows a full consideration of the thermodynamics of RNA folding at varying cation concentrations. It also allows for full folding simulations of the ribozyme, which verify the thermodynamic results of unfolding simulations, and agrees with the observed docking/undocking kinetic behavior from single-molecule work. These atomically detailed folding simulations allow us to observe the tertiary interactions responsible for folding as they are formed, and we model the TSE for docking using extensive p_{fold} analysis and by studying the folding kinetics as a function of tertiary/

secondary interaction strengths. (The division into tertiary/secondary interactions is used both for thermodynamic and kinetic analysis). To compare our simulated results with experimental kinetic data on docking we use the ratio of docking to undocking rates, a quantity that is dimensionless in both Monte Carlo simulation and in experiment. Introducing this method allows a mapping between parameters in our model and experiment which may be generally useful for the interpretation of other Monte Carlo kinetic simulations.

The thermodynamic results predict three major regions of cation/temperature phase space – therefore the topology of the hairpin ribozyme alone dictates its thermodynamic behavior. The kinetic results recapitulate the dynamic docking/undocking behavior of the ribozyme and show a relatively compact TSE with few tertiary contacts formed, in agreement with the experimental model of the TSE. Again, the topology of the ribozyme as encoded in a Go-type model is sufficient to predict the structure of the TSE, a finding that could be useful for identifying the folding pathway for novel RNA structures. Finally, our results shed light on the dynamics of the undocked conformation of the ribozyme, suggesting that high time-resolution fluctuations might be unobserved in existing experiments.

Results

Model and Unfolding Simulations at $\tau/S = 0$

The present model uses a modified Go potential, where interactions are categorized as tertiary or secondary based on the native structure (See Model and Methods for complete description). The strength of different interactions can be scaled by separate tertiary structure interaction strength (τ) and secondary structure strength (S) parameters. For example at $\tau = 0$ and $S = 1$, only secondary interactions contribute to the energy, while tertiary contacts between domains are ignored. Since cations preferentially stabilize tertiary structure, varying the τ/S ratio corresponds roughly to adjusting the ion concentration, but also queries the role of tertiary contacts in kinetic folding to the native state.^{29, 36, 37} In this model we hold $S = 1$ and only vary τ . This model implicitly includes specific cation binding sites *via* the Go-type potential.

The model makes a small set of assumptions that allow us to systematically study the behavior of the ribozyme as a function of τ/S and temperature, and compare these results with measured thermodynamic data. The Go portion of the model assumes that native interactions dominate the folding landscape, and therefore that RNA topology determines the folding pathway. By altering the Go model to include two types of interactions (instead of just one), we assume that tertiary interactions are a combination of attractive and repulsive interactions that are modulated at close range by altering counter-ion concentration. *In effect, our simulations ask the question, what thermodynamic behavior do we predict with these two simple assumptions?* We assess the success of the model by comparing the thermodynamic results to experimentally measured data on thermodynamics and by benchmarking kinetic predictions about the structure of the TSE to experimental measurements.

A model of the two-domain minimal hairpin ribozyme, here called hp2d, was constructed from the crystal structure of the full ribozyme (Figure 1; see Model and Methods). In all subsequent results, the structure hp2d will simply be referred to as “the hairpin ribozyme”. To evaluate the model, unfolding simulations were performed with $\tau/S = 0$ at a variety of temperatures. At low temperatures undocking was observed without substantial dissolution of secondary structure in domain A or B (Figure 2A, B). Undocking here refers to the loss of tertiary structure (Domain A – Domain B interaction) without complete unzipping of either secondary structure domain.

Unfolding via undocking occurs to an intermediate ensemble of states, which consist mostly of a variety of undocked structures without any A–B stacking (Figure 2C) but with intact secondary structure within domains. Structures with stacking between A and B, where the helix continues (with an interruption) via stacking between the ends of domain A and B, also occur (Figure 2D). Brief re-docking events are also observed, for example at ~15,500 in Figure 2A, to a structure very close to the native (Figure 2E shows an overlay of the re-docked structure (blue) with the native hairpin ribozyme (orange)).

Phase Diagram of Hairpin Unfolding

Cation concentration and temperature define a phase diagram for the hairpin ribozyme. At high temperature, one expects all molecules to unfold. At low temperature and high ionic strength a stably docked structure should be observed. From *a priori* considerations it is impossible, however, to make quantitative predictions about the nature of this phase diagram (e.g., where boundaries will lie), or to deduce whether phases other than folded/unfolded will occur. To map out this space, we conducted unfolding simulations of the hairpin ribozyme over a range of temperatures from 1 to 10 and τ/S ratios of 0 to 1.0 (Figure 3). Note that all temperatures are Monte Carlo temperatures, but our calculations suggest that MC temperature of 4.5 corresponds to approximately 315 degrees Kelvin (see Discussion for details).

Comparison of overall structure formation (Figure 3A) with secondary structure alone (Figure 3B) shows at least three distinct regions of the phase diagram. We track the total dRMS at each point in the phase diagram in Figure 3A – this quantity indicates both secondary and tertiary structure formation (low dRMS; white or light gray in the figure) or dissolution (high dRMS; dark gray in the figure). In Figure 3B the sum of the dRMS values for each of domain A and B alone is plotted – this indicates secondary structure formation/dissolution without any tertiary structure contribution. By comparing overall structure (3A) with secondary structure alone (3B) a discrepancy is apparent in the lower-left hand corner (low τ/S and T), where secondary structure is intact but total structure is not (total dRMS is non-zero).

At high temperature, the hairpin unfolds, with high total and intra-domain dRMS. At low temperature and high τ/S the structure is native, with low dRMS for both the entire structure and within each isolated domain. At low temperature and low τ/S , a third phase emerges with intermediate total dRMS but very low intra-domain dRMS. In other words, secondary structure is entirely intact, while the overall structure has non-native dRMS. Trajectories in this region show repeated undocking and docking. Note that domains A and B in isolation unfold at $T \sim 4.5$, with only a very slight preferential stability at lower τ/S (Figure 3B). Indeed, for low τ/S and low T secondary structure is nearly entirely complete, while tertiary structure is partly dissolved.

To more clearly illustrate the three phases we draw a schematic of the unfolded, folded and docking regions in Figure 3C. In the schematic solid lines indicate the midpoint of a transition between two states, while dashed lines indicate the beginning/end points of a transition. It is interesting to note that docking/undocking behavior ranges over τ/S from 0 up to slightly more than 0.3, which gives an estimate for the ratio of tertiary interaction strength to secondary strength in the docking region. τ represents an average over all tertiary interactions, including hydrogen bonds, but also stacking interactions or salt-bridges, and S likewise averages over all interactions within the domains. This phase diagram suggests that, averaged over all contacts, tertiary interactions contribute one-third the enthalpy of secondary interactions. We discuss the implications of this result in the discussion.

Hairpin Folding Trajectories

To verify the equilibrium behavior observed over the phase space of τ/S and temperature via unfolding simulations of the hairpin ribozyme, it is necessary to also find the same results for folding simulations. This is of some computational difficulty, because while unfolding occurs rapidly, folding, even with a Go-like potential, can take much more computational time. To reduce the size of the computational load, we focus on a set of three representative sections on the phase diagram: (1) in the unfolded region, (2) in the docking region and (3) in the folded region (Figure 3C; see Model and Methods for details). In each region we run simulations from the undocked structure long enough to observe a significant number of folding events.

All three regions of the phase diagram observed by unfolding simulation can be recapitulated in folding experiments. In region 1 no folding is observed during the 300,000 step runs. The simulations explore regions of structural space with both secondary and tertiary structure dissolved, which can be seen clearly by tracking Q_{sec} and Q_{tert} , the fractions of native secondary and tertiary contacts, respectively (Figure 4A). In region 2 docking/folding was seen in 47/104 (45%) cases (where folding is defined as attainment of a structure with dRMS less than 0.5 Angstroms by 300,000 steps), confirming that folding is possible, and that docking (to low dRMS and high Q_{tert}) occurs. In this region, secondary structure is largely intact (high Q_{sec}), and the unfolded ensemble makes many forays into nearly-docked structures (Figure 4B) and occasionally docks but does not remain docked (Figure 4B). Finally, in region 3 41/116 (35%) structures fold (dock) within the allotted simulation time of 300,000 MC steps -- it appears that docking occurs slightly less frequently in region 3 than 2. Prior to folding in region 3, the RNA explores the unfolded ensemble with sampling in the vicinity of the docked structure, and after folding the structure does not undock/unfold (Figure 4C) – it remains stable for the duration of all simulations.

Folding Landscape

For subsequent analysis of the Transition State Ensemble (TSE), it is important to understand how many minima characterize the free energy landscape. We histogram data on dRMS (low dRMS indicates folded) and Q_{tert} (high Q_{tert} indicates folded) from 30 folding runs in region 2 (dock/undock; Figure 5A). In the histogram, high density is darker, low density is lighter. There are two major basins of high density: One at very low Q_{tert} with a range of dRMS, the undocked states, and a second at a Q_{tert} of 0.6 to 0.9 with dRMS $< 2 \text{ \AA}$, the docked state. There is a band of low probability density connecting the two states, with a small set of non-productive misfolded states clustered at low dRMS and $Q_{\text{tert}} \sim 0.1$. The folding landscape is therefore primarily two-state, docked and undocked.

To measure the influence of tertiary structure on the TSE, we run sets of folding simulations at varying τ , and measure the undock and dock dwell times to find the docking and undocking rates (Figure 5B). The analysis of dock and undock dwell times is automated, with a “docking” event occurring when the dRMS falls below 0.5 \AA , and “undocking” scored when dRMS rises over 10 Angstroms. Based on the previous analysis (Figure 5A), we interpret the docking and undocking kinetics as a function of τ using a two-state picture.

We consider the hairpin ribozyme docking reaction as a two-state system (Figure 5C). By measuring undock dwell times we determine k_{dock} , which reports on the $\Delta G_{U-\ddagger}$. By measuring the dock dwell times we find k_{undock} , which reports on $\Delta G_{D-\ddagger}$. By varying τ and measuring the relevant rates we can find the effects of altered τ on these free energies, and thus make inferences about the influence of τ on the transition state. We measured undock and dock well times at $\tau = 0.1, 0.15, 0.2$ and 0.25 (Figure 6A shows undock dwell times, B

shows dock dwell times) and fit the histogram of times at each value to a single exponential to get k_{dock} and k_{undock} . In the second set of plots (Figure 6C on a linear scale and D on a logarithmic scale to show detail) we show the dock and undock rates as a function of tertiary strength, from the exponential fits to dwell times. The rates show a slowing of the undock rate, but a nearly constant dock rate. The slowing undock rate implies an increase in $\Delta G_{\text{D-}\ddagger}$, while the constant dock rate indicates that $\Delta G_{\text{U-}\ddagger}$ is constant as a function of τ . We know that the docked (D) state decreases in energy as τ increases – this explains the increase in $\Delta G_{\text{D-}\ddagger}$. Since the undocked (U) state contains very few tertiary interactions, the energy of that state is not affected by changing τ . We conclude that the transition state (\ddagger) contains few tertiary interactions because $\Delta G_{\text{U-}\ddagger}$ is nearly constant as a function of τ .

P_{fold} Calculations

Conclusions about the TS based on undock and dock rate measurements are necessarily coarse and not atomically detailed. To develop a detailed model of the TSE we employ P_{fold} calculations.³⁵ Briefly, we expect structures encountered in time immediately prior to folding to be members of the TSE. During each folding run we save three structures immediately preceding folding, at 10, 50 and 100 steps before folding is detected. This set of structures from all simulations is the putative TSE. We can ask the question, for each structure, is this truly a member of the TSE? If a structure is in the TSE, it should have 50% probability of docking rapidly and 50% probably of undocking – post-TSE structures have close to 100% probability of folding, while pre-TSE structures have close to 0% probability. This probability of folding (calculated over ~40 short simulations) is termed P_{fold}. We calculate P_{fold} for all members of the TSE.

The results of P_{fold} calculations (Figure 7A) for 128 structures are shown. For each structure we calculate Q_{tert} and the dRMS, and then indicate P_{fold} by color. Intermediate P_{fold} values, indicating members of the TSE, are in green. Note that the TSE structures are mostly at Q_{tert} < 0.3 and dRMS < 3 Å, between two regions of post-TSE (blue) and pre-TSE (red) structures. These results show that the TSE is compact (low dRMS) and lacking in most tertiary structure.

To make atomically-detailed predictions about the contacts in the TSE, we calculate simulated phi values. In macromolecular folding experimental phi values indicate the extent of an atom's or functional group's participation in the TSE – a phi of 1 indicates full formation of structure at that position in the TSE, while phi of 0 indicates that the position is entirely unstructured (non-native) in the TSE.³⁴ Here we calculate simulated phi as:

$$\text{Phi}_{\text{sim,tert}} = N_{\text{tertiary contacts, TSE}} / N_{\text{native tertiary contacts}} \quad (1)$$

Where all calculations are over tertiary contacts only (the same calculation with all contacts is less revealing, since it is mostly tertiary contacts that are formed in the TSE to the docked state examined here).³⁸ We include in the TSE all structures with P_{fold} between 0.4 and 0.6, 22 total structures. Using these calculations there are 112 atoms with non-zero phi in the TSE (the full list is not shown; indicated graphically in Figure 7B).

The simulated phi values are uniformly low, again indicating a minimally structured TSE. Nucleotide-level contacts in the TSE are detailed and compared with experimentally-verified contacts in the Discussion.

To visualize the atoms with non-zero phi values we color an image of the native structure by calculated phi value (Figure 7B): Nucleotides with no atoms showing significant phi values are in blue; all nucleotides with phi > 0.05 are colored red or yellow with nucleotides

containing atoms with a phi over 0.2 in red, and nucleotides with maximum atomic phi values lower than 0.2 are in yellow. For comparison we show all tertiary interactions between the two domains in green in Figure 7C.

Timescale of simulated undocked state fluctuations

In single-molecule trajectories of the two-domain hairpin ribozyme the dominant rates of docking and undocking are very similar (0.008 and 0.005 s⁻¹ for docking and undocking respectively).⁹ Therefore the $\tau = 0.25$ condition is most similar to the single-molecule data, with k_{dock} slightly higher than k_{undock} (Figure 6D). Therefore it would be most accurate to compare the timescale of fluctuations in the $\tau = 0.25$ simulations with the single-molecule data reported in work of Zhuang and co-workers.⁹

The simulated folding trajectories show docking and undocking behavior, as discussed above, but they also have large fluctuations in end-to-end distance or dRMS within the undocked conformation (Figure 8A). Here the end-to-end distance is defined as the distance between the FRET pair attachment sites used in the Zhuang study.⁹ To compare fluctuations between the docked and undocked conformation we selected portions of docked and undocked trajectories from a number of simulations and histogram the end-to-end distance values (Figure 8B). The docked conformation (in red, with a detail shown in the inset) has a mean of 37 Angstroms and is tightly distributed. The undocked conformation is widely distributed, with a mean of 78 Angstroms and fluctuations of the distance to near the mean of the docked conformation. Although the undocked state is weighted towards larger distances, it clearly makes excursions to near the docked state. We sought to measure the timescale of these fluctuations towards the docked state.

As an estimate of the predicted timescale of fluctuations within the undocked state, we search for what we call “collapsed” states (as opposed to docked) where the FRET would be greater than 0.5, or where the end-to-end distance is less than 56 Angstroms (for Cy3 and Cy5). To map from Monte Carlo steps to time we will compare the collapse state dwell times with the dock dwell times within the simulation. For example, if the collapse state dwells are 1000 times shorter than the dock state dwells, we will expect that the collapsed state dwells in an experiment should be roughly 1000 times shorter than the (measured) dock state dwell times.

Using this definition of “collapsed” we measure the dwell time in the collapsed state and fit the data to a single exponential with a time constant of 217 MC steps (Figure 8C). From the docking dwell times (Figure 6B lower-right) we find t_{dock} of 121,000 MC steps. Thus in the MC simulation $t_{\text{dock}}/t_{\text{collapse}} = 557$. Experimentally t_{dock} varies from 0.33s to 200s, because there are multiple undocking rates. Using the ratio of 557 this gives an expected collapse state dwell time constant in experiment of 0.6 ms (for fastest undocking) to 0.3 s (for the slowest undocking time constant). The collapse states of FRET > 0.5 with a time constant in the ms may in fact be observable in single-molecule experiments.

Discussion

We have constructed a simple simulated model for the role of cations/tertiary structure strength in RNA folding and have presented a detailed phase diagram for a functional ribozyme, the hairpin ribozyme derived from Tobacco Ringspot Mosaic satellite RNA. We have also shown the first folding simulations of the hairpin ribozyme to the native structure. Using these simulations we have found that the transition state for ribozyme docking has few tertiary contacts, but is relatively compact, in agreement with experimental measurements of docking kinetics. Finally, we predict rapid fluctuations in the undocked

state of the ribozyme which may be observable by low-noise high time-resolution single-molecule microscopy.

Our model is based on two key assumptions: The folding landscape of the RNA is dominated by fold topology as encoded by native interactions (Go model) and tertiary interactions between secondary structure domains are screened by salt in solution. The model predicts a three-state phase diagram for the ribozyme, with folded (docked), unfolded and docking/undocking regions. This behavior may be very general for structured RNAs. A three or four-state phase diagram for an RNA was first observed by Crothers and co-workers in the 1970's for tRNA.³¹ The three-state phase diagram for tRNA in varying magnesium concentrations was confirmed by NMR exchange measurements.³⁹ More recent experiments with the hairpin ribozyme suggest a three-state phase diagram for the minimal form (the same form studied here).^{9,32} The folding simulations from the undocked state recapitulate these phase regions and offer trajectories of tertiary structure formation at atomic-level resolution.

The model presented here for the role of cations in folding is nearly identical to that used by Pande and co-workers¹⁶ for the tRNA^{Phe}, but here we present a systematic study of the variation of the tertiary structure parameter along with temperature, producing a more finely sampled phase diagram for the hairpin RNA. Here, by noting that altering the ratio of tertiary/secondary energies roughly mimics changing ionic concentration, we are able to define a phase diagram for the hairpin ribozyme. This model is based on the assumption that magnesium or sodium ions have little effect on RNA secondary structure while strongly stabilizing tertiary structure.³³ RNA tertiary structure is stabilized primarily by hydrogen bonding between domains, while coulombic repulsion between neighboring backbones tends to destabilize the native state – folding is a balance between these competing energies.³⁰ Increasing cation concentration screens the repulsion between RNA backbones, thus stabilizing the folded conformation. In the supplemental material we present explicit electrostatics calculations of this preferential stabilization of the docked conformation at higher cation concentrations to justify this assumption for the specific case of the hairpin ribozyme.

The influence of cations on RNA structure has been studied in depth, especially by Misra and Draper³⁶, but these calculations use the NLPB equation to calculate electrostatic energies, which is very computationally expensive. The present work attempts to simplify the explicit electrostatic calculations of such work to allow for rapid atomistic simulation of arbitrary RNA conformations -- unfolded, intermediate, and folded. In the absence of much more efficient electrostatics calculations atomistic simulation of RNA folding requires simplified models of electrostatic effects like the Go-type model presented here. This simplified model makes extensive folding and unfolding trajectories possible, and therefore allows us to calculate and verify the phase diagram for the hairpin ribozyme explicitly. It does sacrifice detail in cation binding and therefore may be less realistic for atom-atom contact predictions where the mode of binding (close-range hydrogen bonds versus longer-range electrostatic screening) may be important.

Comparison with experimental thermodynamics

Fluorescence experiments in bulk and at the single-molecule level suggest a three-state phase diagram for the 2WJ version of the hairpin ribozyme, the same “minimal” form studied here. Klostermeier and Millar³² carefully dissected secondary and tertiary structure transitions for the 2WJ and 4WJ (which includes the two naturally occurring C and D domains in addition to A and B (diagram in Figure 1)). They employed simple UV absorbance to measure secondary structure transitions and bulk FRET (with the donor and acceptor attached at the ends of domains A and B, respectively) to measure tertiary

interactions between domains. For the 2WJ they found a single transition in secondary structure as a function of temperature, with a melting temperature of 321° K. In our model secondary structure melts in a single transition at all τ/S (as measured by dRMS within domains A and B; Figure 3B), mimicking the single transition of the 2WJ observed by UV absorbance.

A global fit to a two-state docking model for tertiary structure formation revealed a transition between docked/undocked at temperatures below 315° K to completely unfolded above 315° K. This transition corresponds to the transition between region 2 (docking/undocking) and region 1 (unfolded) in the phase diagram (Figure 3C). Single-molecule work reveals the region 2 – region 3 (folded) transition. Zhuang and co-workers measured the rates (and therefore stability) of the docking/undocking transition as a function of magnesium, and observed a transition to mostly docked ribozymes at high magnesium (above ~100 mM) or very high sodium concentration (above ~1M).⁹ Together these observations demonstrate the existence of a docking/undocking – unfolded transition and a docking/undocking – folded transition. The predicted single-step transition between stably folded (region 3) and unfolded (region 1) at high cation concentration (Figure 3C) has not been directly measured for the hairpin ribozyme. Our results suggest that the ribozyme melts in a concerted single step (combining tertiary undocking and secondary unfolding) at high cation concentrations. This behavior is in fact observed for tRNA: At lower magnesium levels, tertiary structure melts first, followed by various secondary structure elements, while at higher magnesium levels a single melting step is observed at around 333° K.³⁹ The three-state phase diagram of structured RNAs may be relatively generic, occurring in all RNAs with tertiary structure, or at least in those RNAs with a single tightly-interconnected region of tertiary interactions.

The thermodynamic results presented here make a prediction about the ratio of tertiary to secondary interaction energies in biochemically and physiologically relevant conditions. We find that τ/S is greater than 0 but smaller than 0.3 in the “docking” region of the phase diagram, which corresponds to an intermediate range of cations, for example less than ~100 mM magnesium.⁹ In our model τ is identical for all tertiary interactions – it acts effectively as the mean tertiary interaction, including any stacking or hydrophobic interactions, hydrogen bonds and ionic interactions. Similarly, S averages over all secondary interactions – mostly stacking and prototypical Watson-Crick base pairing. It is important to note that hydrogen bond strengths are context-dependent⁴⁰, and that other atom-atom tertiary interactions may depend on ionic strength for their stability (or extent of repulsion, in the case of coulombic interactions).

The model presented here allows for the construction of a phase diagram for RNA folding of the hairpin ribozyme. Remarkably, given the simplicity of the model, it predicts the existence of an intermediate region on the phase diagram corresponding to formed but docking/undocking helices, and makes predictions about the character of transitions between the three different behaviors as a function of temperature and ionic strength. All of the predicted transitions in this phase diagram have been observed for the hairpin ribozyme or for tRNA. The model also predicts the ratio of tertiary to secondary mean interaction strengths in a way that does not involve serial mutations and measurements of $\Delta\Delta G$.

Folding kinetics and unfolded state dynamics

These are the first atomic-level simulations of the folding of the hairpin ribozyme. They are also the first use of any folding simulation to predict the structure of an RNA folding TSE. Previous work from the Pande group on the tRNA^{Phe} had identified intermediates using a Go-type potential very similar to the one used here, but had not modeled the TSEs between the various identified states.¹⁶ Similarly, centroid modeling of RNA folding for tRNA, B-

RNA from the ribosome and other small RNA tertiary structures (but not the hairpin ribozyme modeled here) was able to identify intermediates on the folding pathway but not TSEs.¹⁰ We note that the docked structure of the hairpin ribozyme has also been the subject of simulation work, although these studies do not address the folding mechanism but instead comment on the role of water molecules in catalysis by the ribozyme.^{41, 42, 43}

Single-molecule experimental work has studied the folding kinetics of the minimal form of the hairpin ribozyme simulated here⁹ as well as the four-way junction^{13, 44} and the uncleaved and cleaved versions of the two-way junction ribozyme.⁴⁵ The docking and undocking kinetics depend on the version of the junction studied and on whether or not cleavage has occurred.^{44, 45} Here we focus our simulations on the non-cleaved version of the two-way junction ribozyme studied by Bokinsky and co-workers, and will compare our results to the transition state studies of that work.⁹

The simulated phi values we calculate are uniformly low, indicating a minimally structured TSE. This can be seen graphically in the depiction of P_{fold} as a function of Q_{tert} and dRMS, where the TSE structures (in green, at middle P_{fold}) are mostly at Q_{tert} of 0.2 or lower (Figure 7A). This indicates less than 20% of all tertiary atom-atom contacts formed. The highest phi values, on two atoms, are 0.45, but most phi values are less than 0.2, indicating that no contacts are fully formed in the TSE, and most contacts are only minimally formed. In order to describe the detailed phi results, we briefly summarize the major regions of tertiary structure: The G7:C44 base pair (all numbering using our construct); the ribose zipper (A29/C44 and G30/A43 interactions); and U62 packing (involving U62, G30, U31, A41, A42 and U61).

Non-zero phi values are spread over the interaction surface between the two domains (Figure 7B) but not all tertiary interactions (Figure 7C) have non-zero phi values. We can compare the atomic phi values calculated here with the experimental values from previous work in the Zhuang lab.⁹ Overall, the generally low phi values calculated here agree with the low phi values derived from experiment. In the ribose zipper G30 (G11; nucleotides in parenthesis are numbering from reference 9), the 2' oxygen has phi of 0.3. In the U packing, U31 (U12) has a phi of 0.3 on the 2' oxygen. In the G:C base pair, the experimental phi was calculated by mutating to an A:U pair, with a phi of 0, that is to say, no effect on docking. Thus the overall picture of the TSE as containing few tertiary contacts agrees between simulation and the few carefully measured experimental phi values.

Furthermore, the TSE model emerging from single-molecule experiments is relatively compact.⁹ Zhuang and co-workers conclude, based on magnesium titration combined with electrostatic models of two different possible TS conformations, that the transition state is compact, with the two domains in close proximity. This agrees with the TSE calculated here, where all mid- P_{fold} structures have dRMS less than or equal to 3 Angstroms (Figure 7A).

The Go-like model presented here does surprisingly well at predicting RNA thermodynamics and the TSE of the hairpin ribozyme folding, however there are discrepancies between our kinetic results and the exact folding kinetics as a function of $[\text{Mg}^{2+}]$ and $[\text{Na}^+]$. In our analysis of k_{dock} and k_{undock} as a function of tau only k_{undock} varies. In the work of Zhuang and co-workers only k_{dock} varies as a function of $[\text{Mg}^{2+}]$ while both k_{dock} and k_{undock} vary as a function of $[\text{Na}^+]$. As discussed in the supplemental material our model does not explicitly model electrostatics, and only has short-range tertiary and secondary interactions. We hypothesize that a more complete model with long-range electrostatics shielded by specific- or non-specifically-bound cations would be able to more fully model cation concentration effects on hairpin folding kinetics.

Our analysis of the simulated dynamics of the ribozyme suggests that the undocked conformation is highly dynamic on a rapid timescale (Figure 8). These dynamics have not been directly observed in experimental measurements of the ribozyme. However, such dynamics are consistent with the picture of a compact TSE with low chain entropy emerging from single-molecule experiments. The TSE with the two domains positioned for docking, but with few tertiary contacts, is entropically disfavored, leading to a relatively high free energy barrier to folding and the extensive dynamics within the unfolded state (during the search for the rare TSE configuration) that we observe in simulation.⁹ We hypothesize that motional averaging occurs due to the rapid motion of the domains in the undocked conformation and the relatively slow integration time of images taken of the ribozyme. It remains to be seen whether improvements in single-molecule detection would make even faster, yet low-noise, measurements possible.

In the present study we have performed the first atomistic simulations of the folding of the hairpin ribozyme, and made the first computational predictions of the structure of a TSE for tertiary structure formation in a complex structured RNA. To do this we introduced a simple model of electrostatics within the context of a Go-type model. The model produces a physically realistic phase diagram with three main regions, and produces fluctuations of tertiary structure as observed in single-molecule experiments. The model of the TSE we build based on these calculations is compact but with few tertiary contacts formed, in agreement with single-molecule measurements of the hairpin folding process.

Model & Methods

Our model closely parallels the atomistic protein model of Shimada *et al.* in overall concept, and is a modified version of the RNA model of Nivón and Shakhnovich.¹¹ The latter potential was modified to account for the effects of differential tertiary/secondary interaction strength at varying cation concentrations, but the move set and move acceptance criteria (Metropolis) are identical.

Briefly, we use a heavy atom representation of RNA PDB structures, with a 1.2 Angstrom hard-core radius. The Monte Carlo moves consist of one backbone move followed by one base move, and each composite move is accepted according to the Metropolis criterion. Base moves are rotations about the χ sugar-base bond by a randomly chosen angle at a randomly selected residue. Backbone moves are rotations (simple pivot moves) at one, two, or three, consecutive bases about both the α and ζ Phosphate-Oxygen bonds by randomly chosen angles. This type of backbone move produces both large non-local moves and quasi-local moves, selecting for the former when the polymer is extended and the latter in compact states. Angles for backbone and sidechain moves are taken from a Gaussian distribution with a mean of zero and standard deviation of 4 degrees.

Based on the secondary structure diagram of the RNA in question, the user inputs the boundaries of secondary structures (for the hairpin construct here, domain A constitutes C1 through A33; domain B A34 through A70). During the initial read-in of a structure for folding, atom-atom contacts are categorized as secondary (domain i – domain i) or tertiary (domain i – domain j). In all subsequent energy calculations, the total number of secondary and tertiary atom-atom contacts are counted separately. The energy of a conformation (E_{Go}) is then calculated as:

$$E_{Go} = S * n_s + \tau * n_t \quad (2)$$

Where n_s and n_t are the number of native secondary and tertiary contacts respectively. S is the secondary structure strength parameter, and τ is the tertiary structure parameter. Here we fix $S=1$ and only vary τ .

At $\tau = 0$, no tertiary interactions are stabilized, approximating low cationic concentrations. Higher ionic strength is modeled by $\tau > 0$. Specific cation binding sites are not explicitly modeled, so this construction can only model non-specific interactions of cations, such as magnesium or sodium, with the ribozyme.

This model, with separate tertiary and secondary interactions, captures the fact that extra cations stabilize tertiary interactions by screening of coulombic repulsion between neighboring backbones (reviewed in Misra et al. 2003). RNA tertiary interactions are dominated by hydrogen bonding attraction and coulombic repulsion, mostly from negatively charged phosphates on opposing RNA domains. A primary effect of increasing ion concentration is to add to screening between nearby phosphates, thus decreasing the repulsive energy between tertiary-interacting domains and effectively stabilizing the folded state.

Construction of the Hairpin Ribozyme Model

The structure of the four-way hairpin ribozyme was taken from the crystal structure, PDB ID 1M5K (Figure 1A, B).⁵ In this structure a U1A binding loop was attached at the end of Domain B for ease of crystallization. Note that in the figure only one of the two RNAs in the unit cell is shown, and the U1A protein has been removed. For the simulation, the U1A loop was removed and replaced with the loop structure of loop B from NMR studies (PDB ID 1B36).⁴⁶

The wild-type ribozyme has four domains A–D (Figure 1), but the active site resides in domains A and B, and a construct with C and D cut out is catalytically active.⁷ We simulated the minimal Domain A–B ribozyme which has been studied both in bulk and at the single-molecule level^{7,9} so that the *in silico* predictions would be most directly comparable with experiment. Therefore domains C and D were removed, and the substrate and ribozyme strands (at the end of Domain A) were joined by an inserted loop (loop B from 1B36). The final schematic is shown in Figures 1D and 1E, with the structure in 1C. This structure of the two-domain structure is abbreviated as hp2d to differentiate it from the original four-way junction crystal structure.

The two domains of hp2d and contact types between nucleotides are indicated in Figure 1E. Domain A is in blue and domain B in yellow. Nucleotide-nucleotide contacts scored as secondary structure (within a domain) with energy S are indicated in green (dashes and dots). All contacts between atoms in nucleotides connected by a green mark would be scored with energy S . Nucleotide-nucleotide tertiary contacts between domains, as described in the crystal structure reported by Rupert and Ferre-D'Amare, are indicated in cyan.⁵ All tertiary contacts between atoms in nucleotides connected by a cyan mark (that is, atoms which are in contact but are not in the same domain) would be scored with energy τ .

Hairpin unfolding simulations

To evaluate the simulation methodology, unfolding runs at $\tau = 0$, $S = 1$ were performed for 20,000 MC steps. dRMS, along with the dRMS of sub-regions of domains A and B, were recorded every 100 steps. The total dRMS reports on the combined secondary/tertiary structure of the ribozyme, while the dRMS of an individual domain reports on the secondary structure of that domain alone.

Phase Diagram of Hairpin Unfolding

Unfolding simulations were carried out with τ/S parameter ranging from 0 to 1 (steps of 0.1) and T varying from 1 to 10 (steps of 1.0), for a total of 110 points in the $[\tau/S]/T$ phase diagram. Simulations were as described for preliminary unfolding simulations (above). dRMS and domain dRMS were recorded at all points. dRMS over the entire molecule contains contributions both from tertiary and secondary structure, since both undocking and helix-unfolding will cause deviations from low dRMS. The intra-domain dRMS is calculated separately for four sub-regions of structure within the hairpin, the upper and lower halves of domains A and B. Each intra-domain dRMS is not affected by global changes in structure (such as undocking) but instead only reports on the extent to which that domain is native-like. Three runs were conducted at each $[\tau/S]/T$ combination, for a total of 330 independent unfolding runs, and the average result calculated for the various observables at each point.

Hairpin Folding Simulations from Undocked States

Structures in unfolding simulations with high overall dRMS but native-like secondary structure were isolated and one with the most intact helices (termed UD) was selected for subsequent folding simulation. Folding simulations with the Go-like potential defined by hp2d were carried out for 300,000 MC steps starting from UD. Three regions in the phase space were chosen for study: (1) unfolded region [$T = 8$, $\tau/S = 0.2$], (2) docking/undocking region [$\tau/S = 0.2$, $T = 2$], (3) folded region [$\tau/S = 0.9$, $T = 2$]. Other simulations were also conducted at $\tau/S = 0.2$, $T = 3,4$ and $\tau/S = 0.3$, $T = 2,3,4$, but most results reported are in the three regions described above.

For the more in-depth kinetics measurements folding simulations from undocked were also carried out for 300,000 MC steps at $T=2$ and $\tau/S = 0.1, 0.15, 0.20$ and 0.25 . Dock and undock dwell times were automatically determined from these data by applying a threshold dRMS of 0.5 Angstroms for docking and 10 Angstroms for undocking; the high undocking threshold allows for quick unraveling excursions from docked, which do not result in full undocking. In the experiments of Zhuang and co-workers the A-1 nucleotide is methylated at its 2' hydroxyl, therefore we also carried out similar simulations from the undocked state using an A-1 methylated version of the hp2d structure (Supplemental Material, comparison with methylated ribozyme).

P_{fold} Calculations

To construct a putative TSE we save two structures immediately preceding folding, at 50 and 100 steps before folding is detected, as well as the structure recorded at folding (the nearest multiple-of-ten step), in each folding run. For each of these structures we run ~40 short simulations of 10,000 steps, using the same potential as usual. We calculate P_{fold} by determining how many runs fold (to dRMS < 0.5 Å) and dividing by the total number of runs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Professor Xiaowei Zhuang and Dr. Gregory Bokinsky for helpful discussions on the hairpin ribozyme in general, and the interpretation of docking and undocking rates in particular.

References

1. Lilley DM. Structure, folding and catalysis of the small nucleolytic ribozymes. *Current opinion in structural biology*. 1999; 9:330–8. [PubMed: 10361084]
2. Lilley DM. Structure, folding and mechanisms of ribozymes. *Current opinion in structural biology*. 2005; 15:313–23. [PubMed: 15919196]
3. Hampel A, Tritz R. RNA catalytic properties of the minimum (–)sTRSV sequence. *Biochemistry*. 1989; 28:4929–33. [PubMed: 2765519]
4. Buzayan JM, Gerlach WL, Bruening G. Nonenzymatic Cleavage and Ligation of Rnas Complementary to a Plant-Virus Satellite Rna. *Nature*. 1986; 323:349–353.
5. Rupert PB, Ferre-D'Amare AR. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*. 2001; 410:780–6. [PubMed: 11298439]
6. Fedor MJ. Structure and function of the hairpin ribozyme. *Journal of molecular biology*. 2000; 297:269–91. [PubMed: 10715200]
7. Walter NG, Hampel KJ, Brown KM, Burke JM. Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *The EMBO journal*. 1998; 17:2378–91. [PubMed: 9545249]
8. Zhuang X, Kim H, Pereira MJ, Babcock HP, Walter NG, Chu S. Correlating structural dynamics and function in single ribozyme molecules. *Science*. 2002; 296:1473–6. [PubMed: 12029135]
9. Bokinsky G, Rueda D, Misra VK, Rhodes MM, Gordus A, Babcock HP, Walter NG, Zhuang X. Single-molecule transition-state analysis of RNA folding. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:9302–7. [PubMed: 12869691]
10. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA*. 2008; 14:1164–73. [PubMed: 18456842]
11. Nivón LG, Shakhnovich EI. All-atom Monte Carlo simulation of GCAA RNA folding. *Journal of molecular biology*. 2004; 344:29–45. [PubMed: 15504400]
12. Sorin EJ, Engelhardt MA, Herschlag D, Pande VS. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *Journal of molecular biology*. 2002; 317:493–506. [PubMed: 11955005]
13. Sorin EJ, Rhee YM, Nakatani BJ, Pande VS. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical journal*. 2003; 85:790–803. [PubMed: 12885628]
14. Sorin EJ, Rhee YM, Pande VS. Does water play a structural role in the folding of small nucleic acids? *Biophysical journal*. 2005; 88:2516–24. [PubMed: 15681648]
15. DePaul AJ, Thompson EJ, Patel SS, Haldeman K, Sorin EJ. Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics. *Nucleic acids research*. 2010; 38:4856–67. [PubMed: 20223768]
16. Sorin EJ, Nakatani BJ, Rhee YM, Jayachandran G, Vishal V, Pande VS. Does native state topology determine the RNA folding mechanism? *Journal of molecular biology*. 2004; 337:789–97. [PubMed: 15033351]
17. Go N, Abe H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers*. 1981; 20:991–1011. [PubMed: 7225531]
18. Baker D. A surprising simplicity to protein folding. *Nature*. 2000; 405:39–42. [PubMed: 10811210]
19. Shea JE, Brooks CL 3rd. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annual review of physical chemistry*. 2001; 52:499–535.
20. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of molecular biology*. 1998; 277:985–94. [PubMed: 9545386]
21. Sosnick TR, Pan T. Reduced contact order and RNA folding rates. *Journal of molecular biology*. 2004; 342:1359–65. [PubMed: 15364565]

22. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of molecular biology*. 2001; 308:79–95. [PubMed: 11302709]
23. Shimada J, Shakhnovich EI. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:11175–80. [PubMed: 12165568]
24. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *Journal of molecular biology*. 2000; 298:937–53. [PubMed: 10801360]
25. Li L, Shakhnovich EI. Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:13014–8. [PubMed: 11606790]
26. Hubner IA, Shimada J, Shakhnovich EI. Commitment and nucleation in the protein G transition state. *Journal of molecular biology*. 2004; 336:745–61. [PubMed: 15095985]
27. Hubner IA, Edmonds KA, Shakhnovich EI. Nucleation and the transition state of the SH3 domain. *Journal of molecular biology*. 2005; 349:424–34. [PubMed: 15890206]
28. Fang X, Littrell K, Yang XJ, Henderson SJ, Siefert S, Thiyagarajan P, Pan T, Sosnick TR. Mg²⁺-dependent compaction and folding of yeast tRNAPhe and the catalytic domain of the B. subtilis RNase P RNA determined by small-angle X-ray scattering. *Biochemistry*. 2000; 39:11107–13. [PubMed: 10998249]
29. Misra VK, Draper DE. On the role of magnesium ions in RNA stability. *Biopolymers*. 1998; 48:113–35. [PubMed: 10333741]
30. Draper DE, Grilley D, Soto AM. Ions and RNA folding. *Annual review of biophysics and biomolecular structure*. 2005; 34:221–43.
31. Cole PE, Yang SK, Crothers DM. Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry*. 1972; 11:4358–68. [PubMed: 4562590]
32. Klostermeier D, Millar DP. Helical junctions as determinants for RNA folding: origin of tertiary structure stability of the hairpin ribozyme. *Biochemistry*. 2000; 39:12970–8. [PubMed: 11041862]
33. Misra VK, Draper DE. A thermodynamic framework for Mg²⁺ binding to RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:12456–61. [PubMed: 11675490]
34. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen DE. Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:10426–9. [PubMed: 7937968]
35. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES. On the transition coordinate for protein folding. *Journal of Chemical Physics*. 1998; 108:334–350.
36. Misra VK, Shiman R, Draper DE. A thermodynamic framework for the magnesium-dependent folding of RNA. *Biopolymers*. 2003; 69:118–36. [PubMed: 12717727]
37. Crothers DM, Cole PE, Hilbers CW, Shulman RG. The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *Journal of molecular biology*. 1974; 87:63–88. [PubMed: 4610153]
38. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. *Nature*. 2001; 409:641–5. [PubMed: 11214326]
39. Johnston PD, Redfield AG. Study of transfer ribonucleic acid unfolding by dynamic nuclear magnetic resonance. *Biochemistry*. 1981; 20:3996–4006. [PubMed: 7025889]
40. SantaLucia J Jr, Kierzek R, Turner DH. Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*. 1992; 256:217–9. [PubMed: 1373521]
41. Rhodes MM, Reblova K, Sponer J, Walter NG. Trapped water molecules are essential to structural dynamics and function of a ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:13380–5. [PubMed: 16938834]
42. Park H, Lee S. Role of Solvent Dynamics in Stabilizing the Transition State of RNA Hydrolysis by Hairpin Ribozyme. *Journal of Chemical Theory and Computation*. 2006; 2:858–862.

43. Nam K, Gao J, York DM. Quantum Mechanical/Molecular Mechanical Simulation Study of the Mechanism of Hairpin Ribozyme Catalysis. *Journal of the American Chemical Society*. 2008; 130:4680–4691. [PubMed: 18345664]
44. Nahas MK, Wilson TJ, Hohng S, Jarvie K, Lilley DM, Ha T. Observation of internal cleavage and ligation reactions of a ribozyme. *Nature structural & molecular biology*. 2004; 11:1107–13.
45. Liu S, Bokinsky G, Walter NG, Zhuang X. Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic “fingerprinting”. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:12634–9. [PubMed: 17496145]
46. Butcher SE, Allain FH, Feigon J. Solution structure of the loop B domain from the hairpin ribozyme. *Nature structural biology*. 1999; 6:212–6.

\$watermark-text

\$watermark-text

\$watermark-text

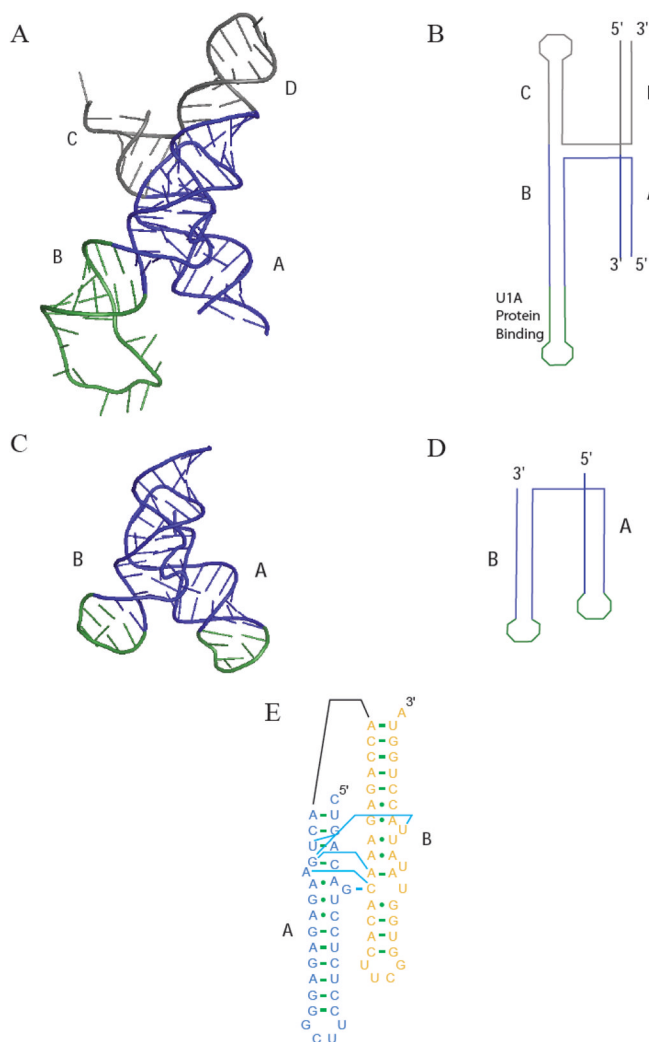


Figure 1. Structure of the four-way junction hairpin ribozyme and the two-way junction model used for simulation. (A) The crystal structure, 1M5K, (absent U1A co-crystallization protein) is shown in cartoon form, with the two-way junction hairpin ribozyme bases colored in blue⁵. Regions removed from this structure in order to build the two-way junction model are the U1A binding loop (green) and domains C and D (gray). Note that the simulation uses all heavy atoms, but for clarity we display only the cartoon model here. (B) A schematic diagram of the secondary structure of the full ribozyme from 1M5K with the same color scheme. (C) The cartoon representation of the model of the two-way junction, hp2d, with loops modeled in from the structure of the isolated Domain B NMR structure(1B36).⁴⁴ The portions of the structure retained from the 1M5K structure are in blue, while loops added from 1B36 are colored in green. (D) A schematic secondary structure diagram of hp2d using the same color scheme. (E) A detailed schematic of hp2d showing all nucleotides, with secondary interactions in green and tertiary in cyan. Within the secondary interactions, dashes indicate canonical Watson-Crick base pairings, while non-canonical pairs are indicated with a dot. Domain A is in blue, domain B in yellow.

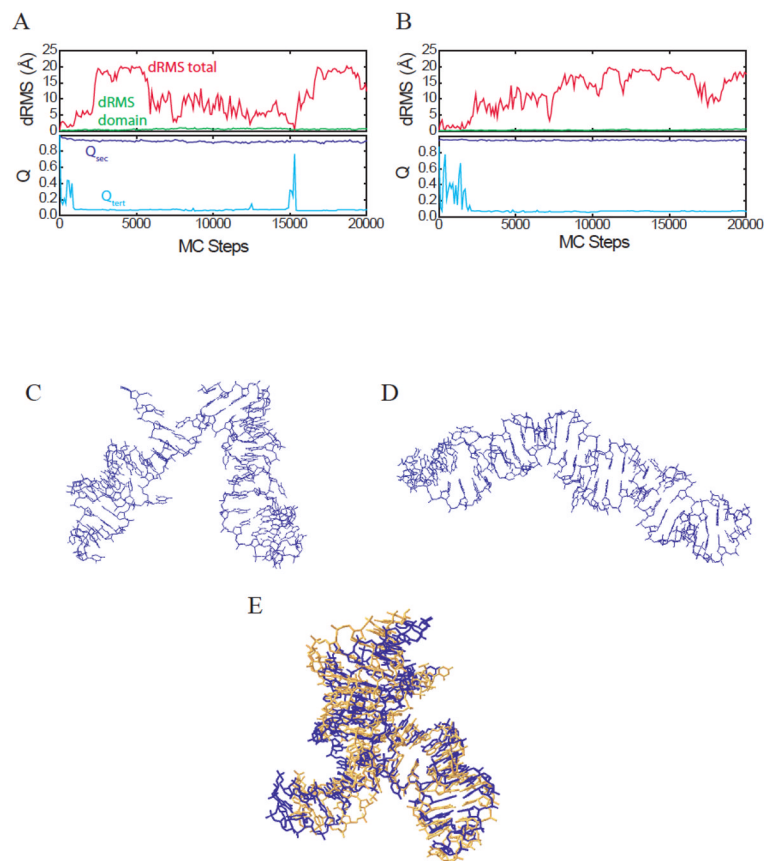


Figure 2.

Sample trajectories and structures from unfolding simulations of the hairpin ribozyme with $\tau/S = 0$. (A) Trajectories of total dRMS (\AA), domain dRMS and fraction of native contacts as a function of number of MC steps show undocking with brief re-docking and (B) undocking without re-docking, but with excursions to an extended “stacked” structure. For A and B total dRMS (red; \AA) and the sum of domain A + domain B dRMS (green; \AA) is shown in the top panel. The fraction of native secondary contacts (Q_{sec} ; dark blue) and native tertiary contacts (Q_{tert} ; light blue) is shown in the bottom panel. (C) Undocked structures occupy a wide swath of structural space away from the stacked structure. (D) An example of the stacked structure. (E) An overlay of a re-docked structure (blue) and the native structure (orange) shows that the re-docked structure is native.

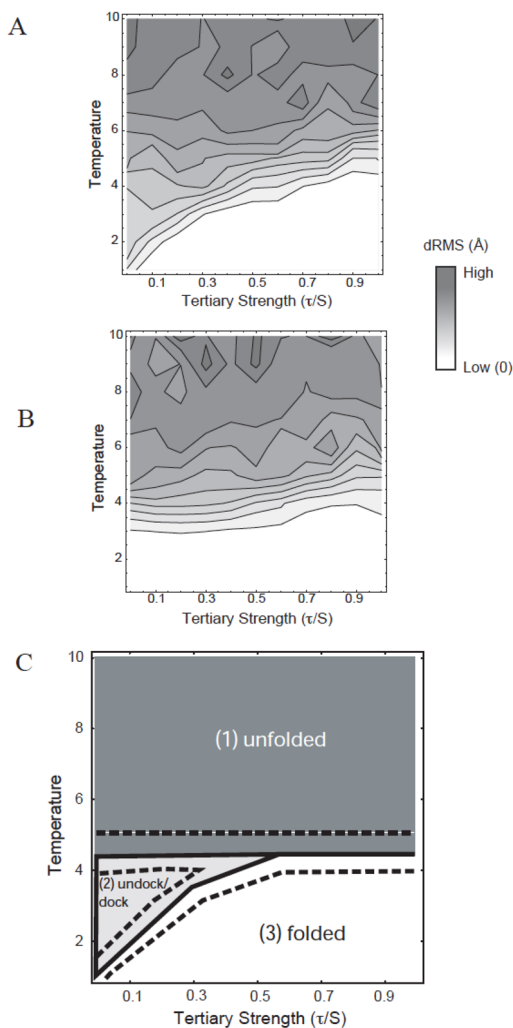


Figure 3.

τ/S and temperature phase diagram of the hairpin ribozyme. The phase diagram was mapped out over $\tau/S = 0.0$ to 1.0 (steps of 0.1) and $T = 1$ to 10 (steps of 1). The ratio τ/S serves as a proxy for experimentally measurable cation concentration in these thermodynamic calculations. Three folding simulations were carried out at each spot in the phase space (A and B). (A) The averaged total dRMS (\AA) at each spot, indicating overall structure formation. (B) The averaged dRMS for domain A plus domain B, calculated separately, indicating secondary structure alone. Higher dRMS (\AA ; less native) is darker. Comparison of the overall dRMS (A) with the isolated region dRMS (B) shows a discrepancy at low T and τ/S , where overall dRMS is intermediate while regional dRMS is very low, indicating loss of tertiary structure with maintenance of secondary structure. By superimposing overall dRMS (A) and domain A + B dRMS (B), one can delineate the three state phase diagram. (C) The regions of phase diagram are shown schematically based on the results of (A and B) in this diagram, showing folded, unfolded and docking regions.

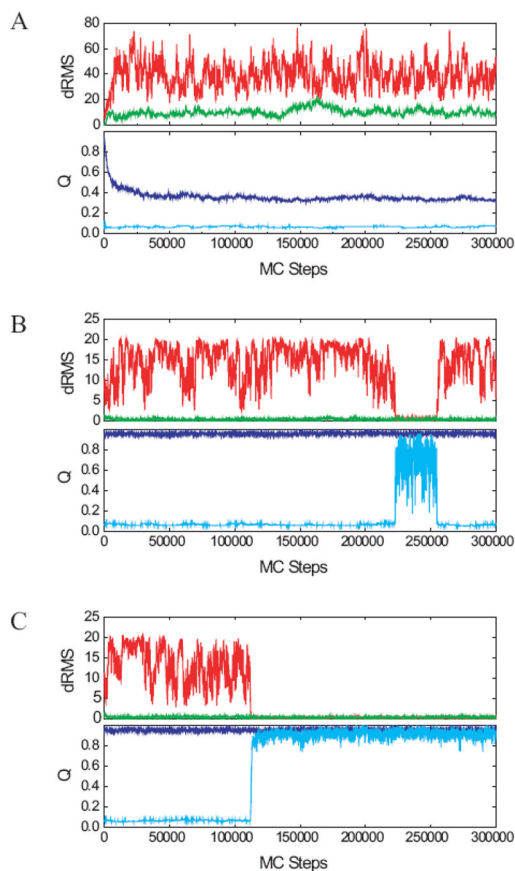


Figure 4.

Folding trajectories in three different regions of the phase diagram (Figure 3C). Folding runs were carried out in three regions of phase space: (A) unfolded (region 1), (B) docking/undocking (region 2) and (C) folded without subsequent undocking (region 3). A structure with intact secondary domains, UD, was extracted from unfolding simulations. Total dRMS (Å; red) and the sum of domain A + domain B dRMS (Å; green) is shown in the top panel. The fraction of native secondary contacts (Q_{sec} ; dark blue) and native tertiary contacts (Q_{tert} ; light blue) is shown in the bottom panel. (A) Folding simulations from the undocked structure, in region 1, unfolded. Note that overall and domain dRMS quickly jump to high values and that Q_{sec} rapidly equilibrates to a low value, indicating unfolding of domains, while Q_{tert} never rises, indicating a lack of any docking. (B) A folding trajectory in region 2 with docking and subsequent undocking. Note in the upper trace that the domain dRMS remains low, while overall dRMS only falls to low values near 230,000 steps, indicating folding. The Q_{sec} remains high throughout, and Q_{tert} only rises during docking. (C) A folding trajectory in region 3 with docking but no subsequent undocking. Undocking is never observed in region 3. The upper trace indicates folding to low overall dRMS around 120,000 steps. The fraction of native tertiary contacts jumps to a stable high value upon folding, while Q_{sec} is not affected.

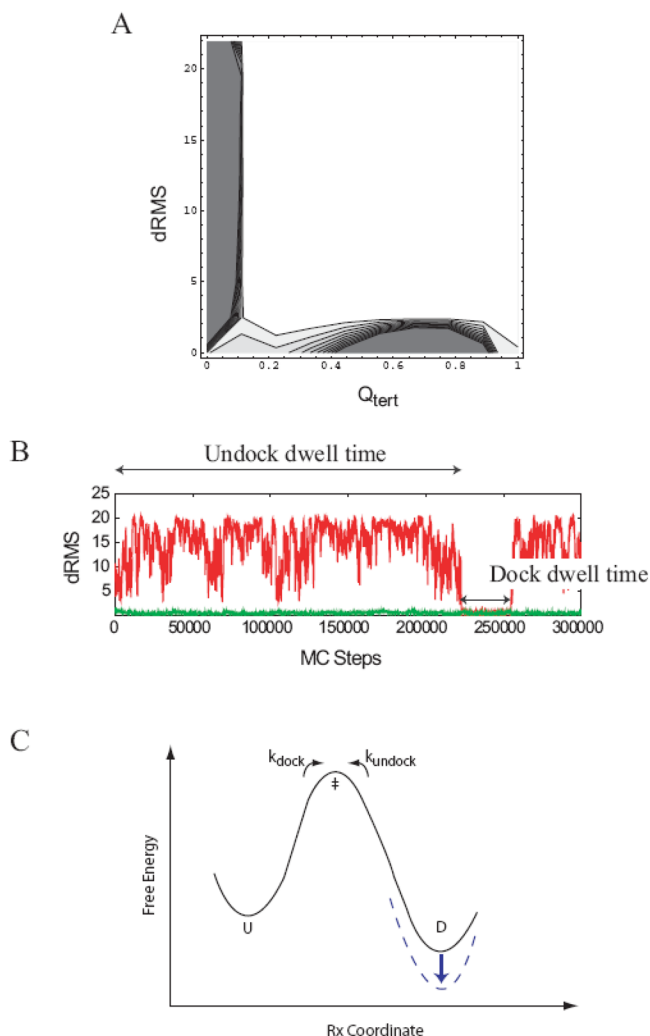
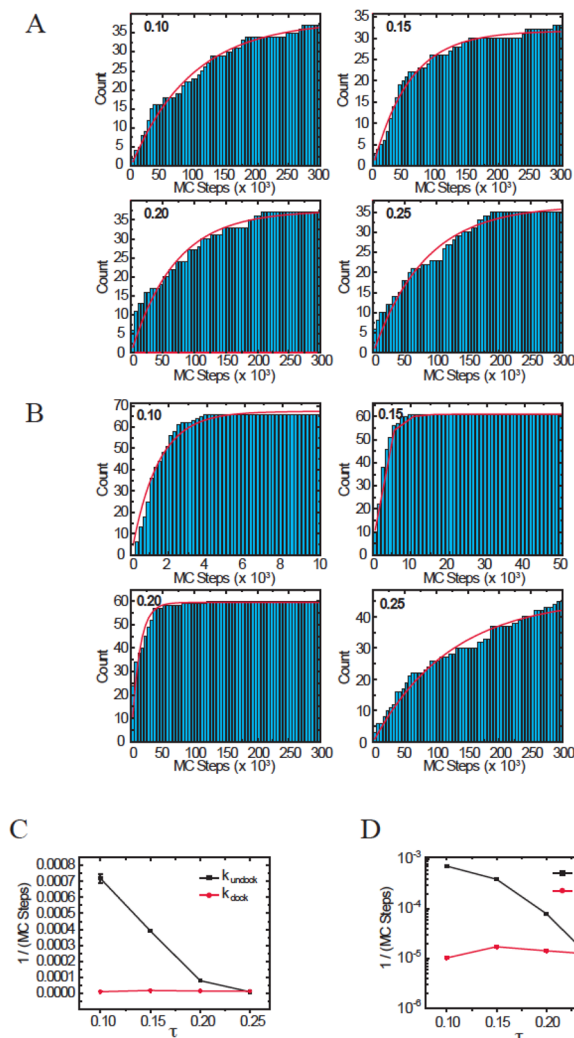


Figure 5. Histogram of the docking/undocking region and schematic of docked/undocked energy landscape. (A) Histogram of dRMS (Å) and Q_{tert} for 30 folding runs in region 2. High density is in dark gray, low density in light gray or white. (B) Undock dwells are defined as the time from the beginning of the simulation until the dRMS (in red) falls under 0.5 Angstroms. The docked state is defined from the end of the undocked state (first excursion of dRMS under 0.5 Å) until the dRMS rises significantly (over 10.0 Å). (C) A schematic of the folding free energy landscape for docking, showing minima at undocked (U) and docked (D). In order to pass from U to D the molecule necessarily traverses a transition state at the free energy maximum. The rate constants k_{dock} and k_{undock} are indicated on the landscape, as well as the influence of stabilization of the D state by increasing τ , or the tertiary interaction strength (dashed blue line).

**Figure 6.**

Undock and dock dwell times as a function of τ , and the docking and undocking rates derived from distributions. Dock and undock dwells were measured over ~ 50 simulations at τ of 0.10 (56), 0.15 (46), 0.20 (54) and 0.25 (44) (numbers in parentheses indicate the number of folding runs at that condition), with the dwells as defined in Figure 5B. All dwells are fit to a single exponential function of the form $y = A * (1 - \exp(-x/t))$, where A is the amplitude and t is the exponential constant. (A) The cumulative undock dwell times with corresponding single exponential fits shown in red. (B) The cumulative dock dwell times with corresponding single exponential fits shown in red. (C) The dock and undock rates ($1/t$) derived from the exponential fits at each τ value. Note that the undock rates vary as a function of τ , becoming slower at higher τ , while the dock rates are not affected by varying τ . (D) The rates ($1/t$) plotted on a logarithmic scale to show detail at the slow rates around $\tau = 0.20$ and 0.25 .

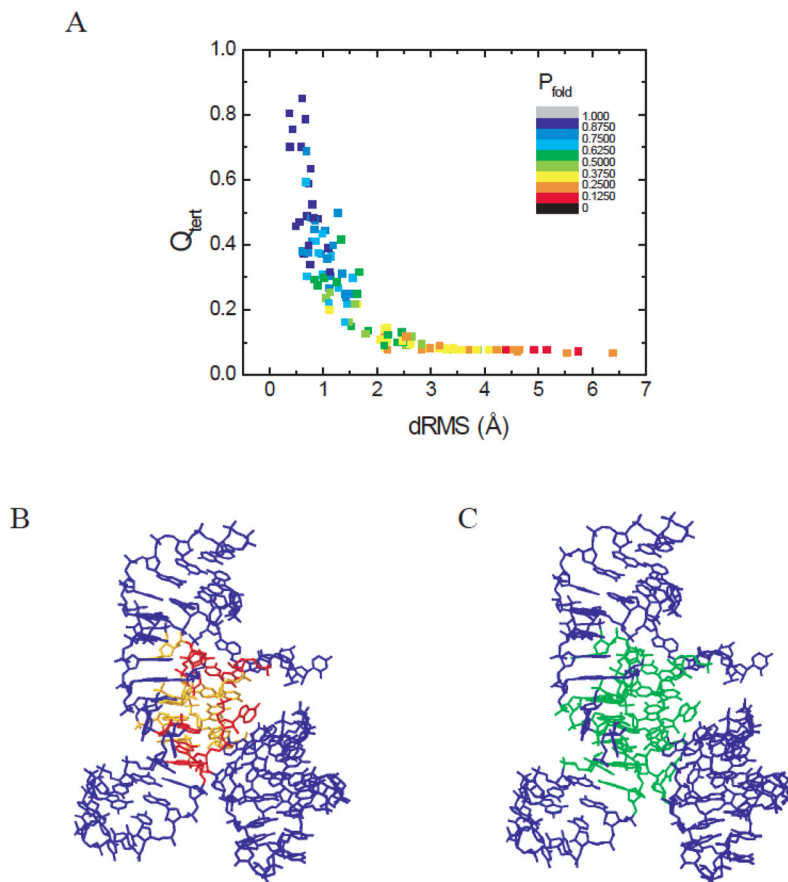


Figure 7.

P_{fold} and simulated phi values for the hairpin ribozyme. (A) The results of P_{fold} calculations for 128 structures are shown. For each structure we calculate Q_{tert} and the dRMS (Å), and then indicate P_{fold} by color as indicated in the legend. Intermediate P_{fold} values, indicating members of the TSE, are in green, while pre-TS structures are in red or yellow and post-TS structures are in blue. (B) The location of non-zero phi value nucleotides on the hairpin ribozyme. Nucleotides are colored according to the atom with maximum phi value. Nucleotides with no phi values above 0.05 are in blue. Those nucleotides with phi > 0.05 are colored red or yellow with nucleotides containing atoms with a phi over 0.2 in red, and nucleotides with maximum atomic phi values lower than 0.2 in yellow. (C) Nucleotides with any atoms showing tertiary contacts with the other domain are shown in green.

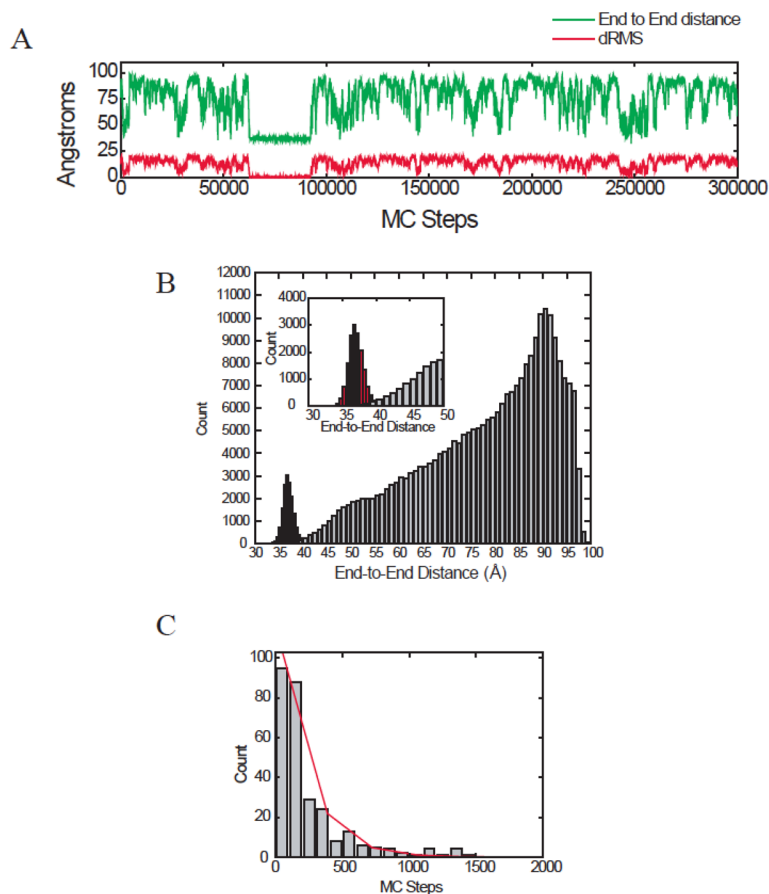


Figure 8.

Conformational distribution and kinetics of the undocked state. (A) A folding trajectory showing both dRMS (Å; red) and end-to-end distance (Å; green) calculated between the two dye attachment points, U17 at the end of domain A and U50 at the end of domain B. Note the docking event (to a relatively fluctuation-free docked conformation) as well as larger fluctuations within the undocked state, in both end-to-end distance and dRMS (B) Histograms of the end-to-end distance within representative sections of docked and undocked conformations (C) Dwell time distribution in the “collapsed” state defined as dRMS < 56 Angstroms (corresponding to the distance of 0.5 FRET for Cy3 and Cy5 dye pair), derived from a section of undocked conformation with no docking. The dwell times in this state are fit to a single exponential with a time constant of 217 MC steps (red line).