doi:10.1016/j.jmb.2004.02.024 J. Mol. Biol. (2004) 337, 789–797

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Does Native State Topology Determine the RNA Folding Mechanism?

Eric J. Sorin[1], Bradley J. Nakatani[1], Young Min Rhee[1] Guha Jayachandran[2], V Vishal[1] and Vijay S. Pande[1,3,4,5]*

[1]*Department of Chemistry Stanford University, Stanford CA 94305-5080, USA*

[2]*Department of Computer Science, Stanford University Stanford, CA 94305-5080 USA*

[3]*Department of Biophysics Stanford University, Stanford CA 94305-5080, USA*

[4]*Department of Structural Biology, Stanford University Stanford, CA 94305-5080 USA*

[5]*Stanford Synchrotron Radiation Laboratory Stanford University Stanford, CA 94305-5080 USA*

**Corresponding author*

Recent studies in protein folding suggest that native state topology plays a dominant role in determining the folding mechanism, yet an analogous statement has not been made for RNA, most likely due to the strong coupling between the ionic environment and conformational energetics that make RNA folding more complex than protein folding. Applying a distributed computing architecture to sample nearly 5000 complete tRNA folding events using a minimalist, atomistic model, we have characterized the role of native topology in tRNA folding dynamics: the simulated bulk folding behavior predicts well the experimentally observed folding mechanism. In contrast, single-molecule folding events display multiple discrete folding transitions and compose a largely diverse, heterogeneous dynamic ensemble. This both supports an emerging view of heterogeneous folding dynamics at the microscopic level and highlights the need for single-molecule experiments and both single-molecule and bulk simulations in interpreting bulk experimental measurements.

Keywords: RNA folding; bulk kinetics; mechanism; distributed computing; biasing potential

Introduction

The study of biological self-assembly is at the forefront of modern biophysics, with many laboratories around the world focused on answering the question "How do biopolymers adopt biologically active native folds?" In recent years, protein folding studies have put forth a new "topology-driven" view of folding by recognizing correlations between folding mechanism and topology of the native state.[1] For instance, the laboratories of Goddard and Plaxco have put forth several reports connecting the folding rates of small proteins with the topological character of those proteins,[2–5] and comparisons between topology and unfolding pathways have been made.[6,7] Ferrara and Caflisch studied the folding of peptides forming beta structure and concluded that the folding free energy landscape is determined by the topology of the native state, whereas specific atomic interactions determine the pathway(s) followed,[8] and our recent comparison between protein and RNA hairpin structures supports this line of reasoning.[9] This is not surprising when one considers that the folding transition states for proteins of similar shape were shown to be insensitive to significant sequence mutations.[10–12]

Here, we attempt to directly characterize the contribution of native state topology to the RNA folding mechanism at both the single-molecule and bulk levels. Because bulk experiments inherently include ensemble-averaged observables, with distributions around those averages expected within a given system, single-molecule studies complement bulk studies by allowing observation

Abbreviations used: RMSD, root-mean-squared deviations; LJ, Lennard–Jones.

E-mail address of the corresponding author: pande@stanford.edu

of dynamics that are masked by such ensemble-averaging.[13–15] Yet differences between single-molecule and bulk folding behaviors can be profound, as detailed herein.

In the past, the ability of simulation to predict bulk system dynamical properties has been hindered by the limited number of events that one could simulate. Using a distributed computing architecture[16] to overcome this barrier, we have collected nearly 5000 complete folding events for the 76 nucleotide tRNA topology, which is significantly more complex than the small, fast-folding systems commonly examined in folding simulations.[9,17,18] This degree of sampling represents ~500 CPU years of work conducted on ~10,000 CPUs within our global computing network†.

Interestingly, topological considerations have only been applied to the smallest of RNAs,[9] likely because the folding of functional RNAs is known to be more complex than that of small proteins due to the coupling of RNA conformational energetics with the predominantly ionic environment.[19] For instance, it has been demonstrated that the identity and concentration of ions present can alter the folding and unfolding dynamics exhibited by the tRNA topology,[20–22] further complicating our understanding of RNA folding.

To directly assess the relationship between native topology and RNA folding mechanism, we employ a minimalist model similar to the Gō-like models previously used to study protein folding.[15,23–28] We improve upon these models by developing a folding potential that: (i) treats the polymer in atomic detail; (ii) introduces attractive non-native interaction energies; and (iii) includes ion-dependent effects on the folding dynamics. These features allow for a much more accurate approximation of the true energetics than the coarse-grained models commonly used to study the dynamics of large biomolecules,[28] while inherently maintaining the polymeric entropy expected from atomistic models, thereby allowing a causal relationship to be inferred between changes in the energetics of the model and changes in the resulting dynamics.

Striving for simplicity, we consider the most elementary model of the ionic effect in RNA folding, which assumes that monovalent and divalent cations stabilize secondary and tertiary structure, respectively. We propose that such a simple model is adequate to capture the essential folding dynamics observed in experiments under varying ionic conditions: our model employs two parameters, ε_2 and ε_3, which specify the energetic benefit of native secondary and tertiary interactions, respectively, coupled only in their simultaneous use in thermal calibrations of the model.

We note that while the use of Gō-like potentials to extract precise folding dynamics is questionable at best, the simplified biasing potential used here presents an ideal methodology to probe the relationship between topology and folding mechanism: the folding potential is built on information from the native tRNA fold. If the relationship between topology and folding mechanism is negligible, our results should show significant disagreement with experimental observations. If, on the other hand, this relationship is significant, the agreement with experiment should also be significant.

Indeed, we show below that the bulk folding behavior predicted by this simple, atomistic, minimalist model predicts well the experimentally observed bulk folding mechanism, suggesting that topology does in fact contribute to the bulk folding mechanism. Furthermore, massive sampling has, for the first time, allowed direct comparison of simulated ensemble kinetics to single-molecule folding simulations, illustrating a tremendous heterogeneity inherent to individual folding events that collectively contribute to a topologically driven bulk folding mechanism. We discuss below the general results seen in single-molecule folding trajectories, followed by analysis of kinetic and mechanistic observations from an ensemble-averaged perspective, as well as the role of polymeric entropy in folding. Caveats of the model are then discussed in Methods.

Results

Diverse pathways in tRNA folding

The native fold of tRNA can be described in terms of nine "substructures" (Figure 1(a)). The well-defined secondary structure is composed of four helix-stem regions (S_1, S_2, S_3, S_4), while the tertiary structure consists of a set of five regions of long-range contact (T_{12}, T_{13}, T_{14}, T_{23}, T_{24}), where the dual subscripts denote the two helix-stems in contact for a given tertiary region. Formation of these substructures in simulated single-molecule folding events occur as a series of numerous, discrete transitions in the fraction of native contacts, Q. To quantitatively characterize the observed pathways, the relative folding times for these nine substructures were calculated. In many cases, multiple substructures folded in the same time range. To unambiguously determine the substructural folding order in a given trajectory, the variance in the fraction of native contacts within each substructure throughout that trajectory, Var(Q), was calculated, and the relative folding time of each substructure was defined as the temporal point at which Var(Q) was a maximum (i.e. the point at which Q was increasing most rapidly).

Pearson correlations (defined as the covariance of two sets divided by the product of their standard deviations) between folding times for each pair of substructures within the ensemble of folding events were then evaluated, and the resulting

† Folding@Home, http://folding.stanford.edu

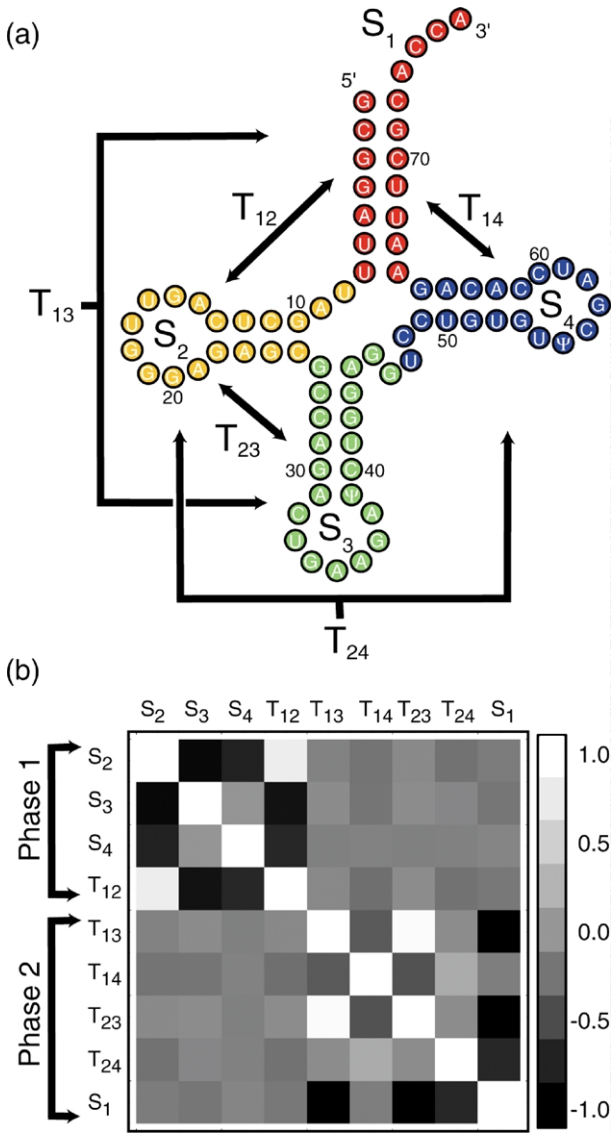

Figure 1. tRNA substructure depiction and folding time correlations. (a) tRNA substructures: red, S_1 acceptor stem; orange, S_2 D-stem; green, S_3 anti-codon stem; blue, S_4 T-stem; regions of tertiary contact are noted with arrows, where T_{ij} indicates tertiary interactions between the S_i and S_j stems. (b) Relative folding time (Pearson) correlations between each substructural pair within the ensemble of 2592 folding trajectories.

matrix (Figure 1(b)) shows a distinct separation between two phases during folding. These uncorrelated groups are referred to below as phase $1 = \{S_2, S_3, S_4, T_{12}\}$ and phase $2 = \{T_{13}, T_{14}, T_{23}, T_{24}, S_1\}$. Each folding trajectory was then assigned to a specific folding pathway, denoted by the sequence of substructural folding events that occurred in that specific simulation.

In an effort to simplify the representation of the many pathways observed, folding simulations were grouped into three classes. While the initial events in the folding process were similar for all folders (predominantly stem formation), the final stages of folding exhibited a much larger variation. For this reason, the ordering of the substructural folding events within phase 2 was used to classify folding trajectories. The observed statistical weighting of folding pathways is detailed for each folding class in Table 1. For brevity, pathways that contributed less than 1% of the folding ensemble have been removed, and the tabulated pathways account for approximately 90% of observed folders.

Table 1. Weighting of tRNA folding pathways

Class	Weight (%)	Pathway
I	36.3	S_4 [S_2,T_{12}] S_3 [T_{13},T_{23}] [T_{14},T_{24}] S_1
	10.6	S_4 S_3 [S_2,T_{12}] [T_{13},T_{23}] [T_{14},T_{24}] S_1
	4.5	[S_2, T_{12}] S_4 S_3 [T_{13},T_{23}] [T_{14},T_{24}] S_1
	1.7	S_4 [S_2,T_{12}] S_3 T_{24} [T_{13},T_{23}] T_{14} S_1
II	13.0	S_4 [S_2,T_{12}] S_3 S_1 [T_{13},T_{23}] [T_{14},T_{24}]
	3.4	S_4 S_3 [S_2,T_{12}] S_1 [T_{13},T_{23}] [T_{14},T_{24}]
	2.2	S_4 [S_2,T_{12}] S_3 S_1 T_{14} [T_{13},T_{23}] T_{24}
	1.5	S_4 [S_2,T_{12}] S_1 S_3 [T_{13}, T_{23}] [T_{14},T_{24}]
	1.4	[S_2,T_{12}] S_4 S_3 S_1 [T_{13},T_{23}] [T_{14},T_{24}]
III	8.1	S_4 [S_2,T_{12}] S_3 [T_{13},T_{23}] S_1 [T_{14},T_{24}]
	2.4	S_4 S_3 [S_2,T_{12}] [T_{13},T_{23}] S_1 [T_{14},T_{24}]
	2.4	S_4 S_3 [S_2,T_{12}] [T_{13},T_{23}] S_1 [T_{14},T_{24}]
	1.6	S_4 [S_2,T_{12}] S_3 [T_{13},T_{23}] T_{14} S_1 T_{24}

Although this sampling of pathways cannot quantitatively capture equilibrium thermodynamic behavior, insight into the qualitative nature of the folding landscape (e.g. locations of intermediates and free energy barriers) can be gained by considering the relative probabilities of each microstate present in the simulated ensemble along a minimal number of folding parameters (the fraction of native secondary and tertiary contacts, Q_2 and Q_3, respectively). The log of the conformational probabilities on this simplified, two-dimensional surface is plotted in Figure 2(a). It is evident that each class of folding pathways predicts multiple intermediates in the folding process, observed as the previously mentioned numerous, discrete steps in single-molecule folding trajectories. A graphical representation of the most statistically predominant pathway observed is shown in Figure 2(b), and an animation of this simulation is available for viewing online†.

Ensemble-averaged kinetics

We find significant differences in comparing ensemble averaged dynamic properties to single-molecule trajectories. As shown in Figure 3(a), again using the fraction of native contacts as the primary folding parameter, several distinct populations are present throughout the ensemble of folding events. With the exception of the fully extended chain (E, a physically irrelevant state used to start simulations without biasing the resulting folding pathways), these states qualitatively

† http://folding.stanford.edu/tRNA/

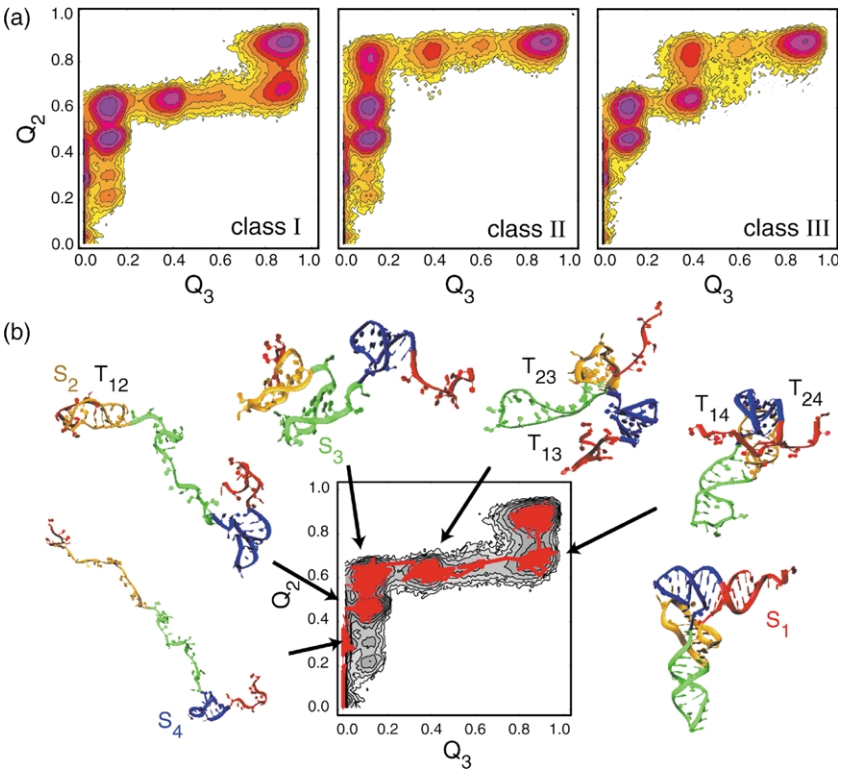

Figure 2. Statistical energy profiles and the predominant folding pathway. (a) Locations of free energy barriers and minima on the simplified two-dimensional folding surface, where folding classes were defined by the order of substructural formation in the second phase of folding. (b) Graphical representation of the most predominant folding pathway in our simulations with substructures color-coded as in Figure 1(a). Labels indicate which substructures have folded or are folding in each frame. The relevant statistical energy landscape is shown in grayscale with the trajectory overlaid in red.

agree with experimentally observed populations described by Sosnick and co-workers,[20,29] who differentiate between multiple unfolded states (U′ and U in the presence and absence of 4 M urea, respectively) and intermediates (I_{Na} denotes the bulk intermediate observed in high sodium concentrations and I_{Mg} is the bulk intermediate observed in the presence of magnesium).

The distributions of native contact formation $P_{fold}(Q)$ *versus* the number of native contacts (Q) for each substructure are plotted in Figure 3(b), with secondary substructures (upper panel) color-coded as in Figure 1(a), and tertiary substructures (lower panel) scaled from white (T_{12}) to black (T_{24}). Early events in the folding process (when Q is small) include formation of the three non-terminal helices, alongside the T_{12} tertiary contact (phase 1). Long-range tertiary contacts $\{T_{13}, T_{14}, T_{23}, T_{24}\}$ then form in any variety of temporal sequences after phase 1 has completed, with the formation of the terminal helix-stem (S_1, red) either preceding or succeeding collapse and formation of tertiary structure, as described experimentally.[20]

The distribution of the radius of gyration R_g is shown in Figure 3(c), and shows quantitative agreement with experimentally measured molecular sizes attained for collapsed states:[29] the I_{Mg} and N ensembles have $\langle R_g \rangle = 31.8(\pm 3.4)$ Å and $\langle R_g \rangle = 26.2(\pm 1.3)$ Å, respectively. These values are only slightly larger than those reported by Fang *et al*. due to the added polymeric flexibility of our model (see Methods). While the quantitative agreement for the native state R_g is fortuitous (based on the construction of the model), the quantitative agreement observed for I_{Mg} is a striking prediction of the character of the intermediate from a model based solely on native topology information. The distribution of all-atom root-mean-squared deviations (RMSD) from the native structure is also shown (inset).

The folding mechanism predicted from a bulk perspective (which is dependent on ion identity and concentration) is shown in Figure 3(d). From the unfolded state, the I_{Mg} (full formation of secondary structure) and I_{Na} (formation of the three non-terminal helix-stems) intermediates are both possible. From I_{Mg}, the I_{Na} intermediate can be formed, with phase 2 tertiary contacts forming last. Alternatively, folding may include formation of tertiary contacts followed by zipping of the terminal S_1 helix-stem.

To illustrate how a three-state ($U \rightarrow I \rightarrow N$) observation could come from a large ensemble of individual folding events, each exhibiting numerous discrete transitions, Figure 4 depicts the ensemble-averaged fraction of native contacts $\langle Q(t) \rangle$ alongside a single, randomly chosen trajectory (Figure 4(a)). Figure 4(b) demonstrates how the discrete nature of folding is lost when considering only ten randomly chosen individual folding events in the averaging. With more than 2500 trajectories (Figure 4(c)), all sense of discrete, stepwise folding is lost, and the $\langle Q(t) \rangle$ curve becomes a smooth function of time. Still, the averaging of only ten independent folding trajectories is more representative of a single trajectory than of the ensemble as a whole, and only after including ~500 or more events (data not shown) does the $\langle Q(t) \rangle$ curve approximate that of the entire ensemble.

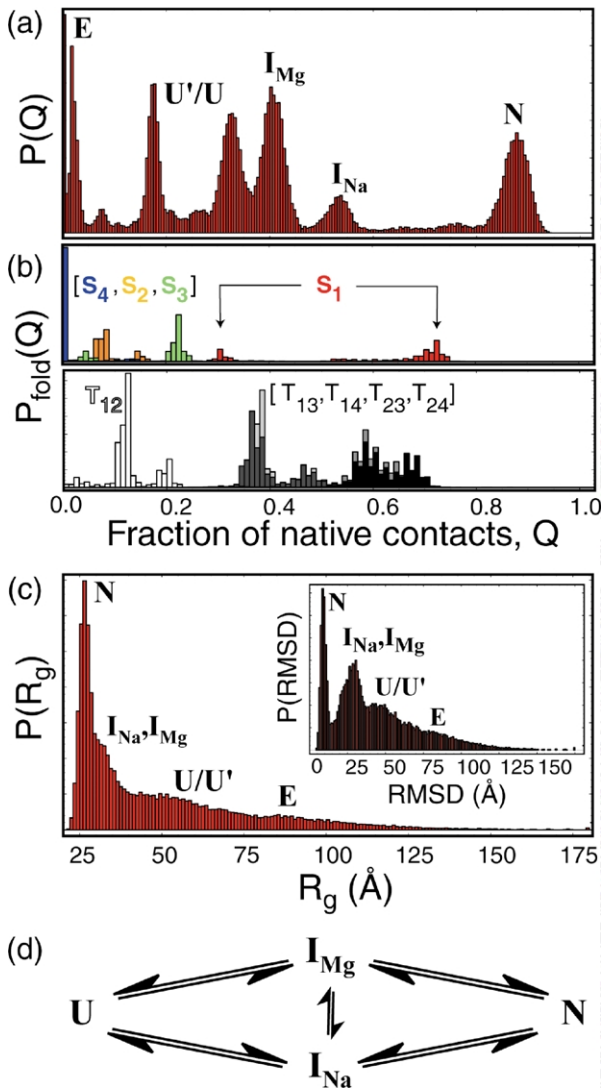

Figure 3. Ensemble-averaged properties during tRNA folding. (a) Detected states in the bulk folding process. (b) Distributions of relative folding times for each substructure, with secondary structures color-coded as in Figure 1(a) (upper panel) and tertiary structures scaled from white to black (lower panel). (c) R_g and RMSD (inset) distributions for the observed states. (d) Predicted ion-dependent bulk folding mechanism.

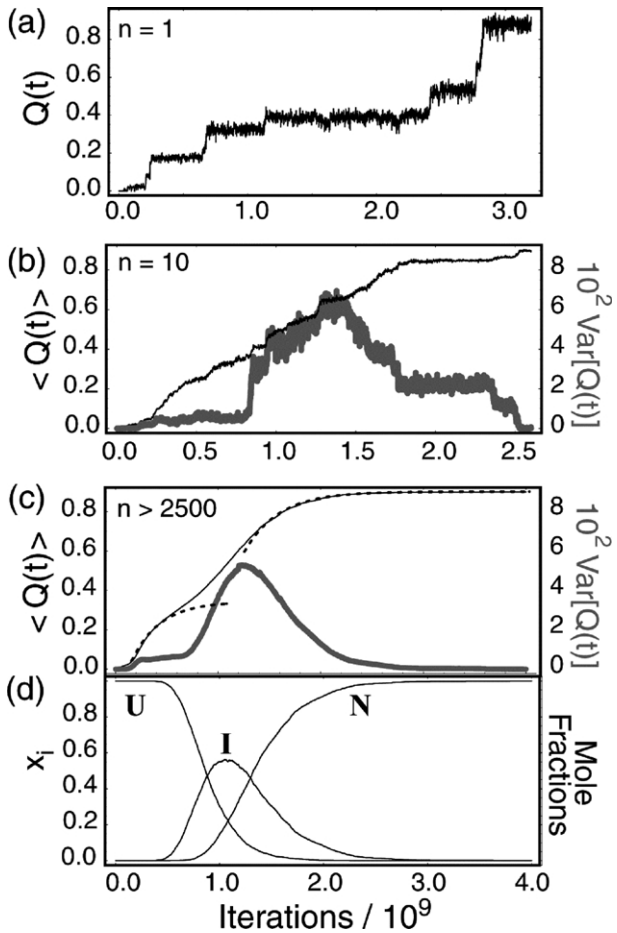

Figure 4. Ensemble-averaged kinetics. The fraction of native contacts for a single folding trajectory (a) shown next to the ensemble-averaged fraction of native contacts for ten randomly chosen folding simulations (b) and all 2592 folding events (c). $\langle Q(t) \rangle$ approaches a smooth function of time with two distinct, sequential exponential phases (fitted as dotted curves) in (c). In (b) and (c), Var[$Q(t)$] is shown as a thick gray line. (d) The time-dependent mole fractions for the U, I, and N states.

It is convenient to consider the formation of each of the nine independent substructures within the ensemble, starting from the extended polymer, as a stochastic (Poisson) process with a given folding rate λ_n. This is particularly the case for simple helix formation (with no bulges or other internal structure), which has been shown to be a two-state process.[9,30,31] Such processes are additive such that the sum of n independent Poisson processes is itself a Poisson process with a total rate equal to the sum of the n individual rates. Thus, if folding occurred simply as simultaneous formation of all nine substructures, one would expect the total number of native contacts Q to increase exponentially in time, with a total rate equal to the sum of the rates for each substructure.

In contrast to the simple, exponentially increasing $\langle Q(t) \rangle$ described above, two such phases are apparent in Figure 4(c). Because phase 2 begins only after a portion of individual trajectories have completed phase 1, the two transitions in the $\langle Q(t) \rangle$ curve were fit individually to single-exponential processes (dotted curves in Figure 4(c)), resulting in R^2 values of 0.999 and 0.998, respectively. Were there not a significant overlap between the two phases (i.e. if all trajectories completed phase 1 prior to any of them starting phase 2), these fits would be expected to fully predict $\langle Q(t) \rangle$, which would then exhibit a cusp at the crossover point between the two fits. As the resulting rate of phase 1 is greater than that for phase 2, the buildup of a single observable intermediate between the phases is predicted.

To consider the dynamic populations present from a chemical kinetics perspective, Figure 4(d) shows the evolution of ensemble mole fractions of

unfolded U, intermediate I, and native state N conformations. The I state was defined (from statistical populations in Figure 2(a)) as having formed ~60% and ~10% of native secondary and tertiary contacts, respectively, and exhibits a concentration maximum that coincides with the crossover point of the individual fits in (c). The native state curve is well fit by the proper biexponential relationship for the sequential, irreversible reaction U → I → N with an R^2 value of 0.999.

Pathways for overcoming entropy

A strength of atomistic, minimalist models is their utility in elucidating the role of polymeric entropy in the folding dynamics.[27] It is thus interesting to consider the balance of interaction energy and polymeric entropy in tRNA folding. We stress that in this minimalist model the energetics are defined by the native tRNA topology or, more specifically, by the "stability ratio," $\varepsilon_2/\varepsilon_3$. The folding mechanisms for extreme values of the stability ratio are obvious: for $\varepsilon_2/\varepsilon_3 \gg 1$ secondary structure will form preferentially, and for $\varepsilon_2/\varepsilon_3 \ll 1$, tertiary structure would be preferred. Simulation can play an important role in characterizing the experimentally relevant intermediate regime, where one must consider the delicate balance between the energetic benefits and entropic penalties of forming native structure.

The variation of $\varepsilon_2/\varepsilon_3$ implies the variation of ionic conditions: calibration of our model to experiment suggests that for $\varepsilon_2/\varepsilon_3 = 1$ (Figure 5, top panels) we are in a regime analogous to magnesium-mediated folding, whereby three of the four helix-stem elements are structured in the early stages of folding (observed experimentally as a preference for I_{Mg}), with S_1 forming late in the folding process. In contrast, when secondary structure is significantly more favorable than tertiary structure, $\varepsilon_2/\varepsilon_3 = 3$, our model simulates a regime analogous to higher sodium concentrations in the absence of magnesium, as shown in Figure 5 (lower panels). In this case, the formation of the terminal stem-helix precedes the formation of phase 2 tertiary contacts (I_{Na}). We stress that this finding is in good agreement with the rather non-canonical experimental results reported by Shelton *et al.*,[20] in which the simple cloverleaf secondary structure is not a necessary intermediate on the tRNA folding pathway.

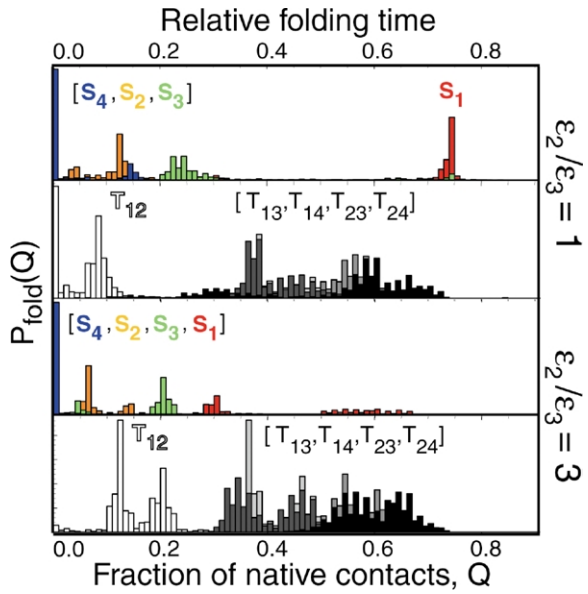

Figure 5. Ion-dependent dynamics. As in Figure 3(b), the distributions of relative substructural folding times are shown for both perturbed values of the stability ratio $\varepsilon_2/\varepsilon_3$.

Like the formation of nucleic acid hairpins, the folding of the terminal RNA helix-stem appears to be cooperative due to a balance between the energetics of interaction and the polymeric entropy loss of structure formation. Indeed, it has been previously suggested that the polymeric entropy of biomolecules can be sufficient to lead to cooperative folding.[32] However, one would expect a significant entropic penalty for bringing together the two ends of the 76 nucleotide tRNA, and surmounting this barrier is apparently accomplished in one of the two experimentally observed ways.

Increasing the stability ratio results in greater secondary structure stability, and the enthalpy of base-pairing dominates the conformational free energy, thus favoring early zipping of the terminal stem (I_{Na}). On the other hand, decreasing the stability ratio implies a more stable tertiary structure, and the entropic barrier to terminal zipping dominates, resulting in unpaired S_1 strands (I_{Mg}) until late in the folding process (after long-range collapse has occurred). We thus hypothesize that this entropic barrier is responsible for the distinct ion-dependent dynamics observed in tRNA folding, with counterions tuning the relative stability of the intermediates and, in effect, "tipping the scale" in favor of one mechanism over another. To be sure, our simulations show that even a somewhat subtle change in the stability ratio can lead to a qualitative change in the folding behavior.

Discussion and Implications

Using massively distributed computational sampling and a minimalist, atomistic model with implicit solvation and counterion effects, we have shown that bulk non-equilibrium reaction dynamics can be extracted from many single-molecule events without the need for extrapolating bulk behavior from a small number of simulations. While our ensemble-averaging includes far fewer events than the number studied in standard bulk experiments, we have observed convergence to an ensemble signal from ~2500 individual folding events, and have identified the lower bound for attaining the ensemble signal to be ~500 individual trajectories for a molecule of this size and complexity.

Our results indicate that native tRNA topology serves as a dominant predictor of the bulk folding mechanism. To be sure, a biasing potential built from native state information that includes non-native interaction energetics recovers the known three-state behavior reported previously and well predicts the character of known environmentally dependent intermediates, as specified by both molecular size and substructure formation.[20] In contrast to the ensemble-averaged behavior, great diversity in single-molecule folding pathways has been observed. This is not surprising when considered in the context of contemporary folding theories, which identify native folds as energetic "traps" and portray the polymeric motion that leads to trapping in low-energy states as predominantly stochastic in nature.

Still, it remains a common practice to interpret bulk signals from various experimental sources as representing the "true" dynamics on a microscopic level, and it has only recently been put forth that (unobservable) intermediates are likely present even in the most simple two-state folders.[33] The sampling reported herein thus serves as an example of the distinction that should be drawn between single-molecule (microscopic) and bulk (macroscopic) probes, and highlights the need for both single-molecule experiments and simulation, both single-molecule and bulk, in the interpretation of bulk measurements. Indeed, it will be intriguing to see this complexity revealed by single-molecule experiments of tRNA folding.

Methods

Construction of the minimalist model

To construct a minimalist model for RNA folding, we have expanded upon previous minimalist models of protein folding[15,25,28,34,35] by including non-native long-range interaction energy terms (adding energetic frustration to the folding landscape) and by considering the role of solvated counterions in the folding process. We define native contacts as atomic pairs separated by three or more residues and within 6.0 Å of one another in the native structure (based on $tRNA^{phe}$, PDB ID 6tna). These interatomic interactions are modeled using different classes of long-range Lennard–Jones (LJ) interactions: native contact pairs within the same helix-stem region (secondary contacts) are assigned energy ε_2 (initially ~ten times greater than non-native interaction energies), while all other native pairs (tertiary contacts) are assigned LJ energy ε_3.

An initial stability ratio of $\varepsilon_2/\varepsilon_3 = 2$ was used, thereby assigning twice the energetic benefit to native secondary contacts as that assigned to native tertiary contacts, and a total of 2592 complete folding events were simulated. To gain insight into the distinct kinetics reported for differing ionic conditions,[20] over 1000 complete folding events were also collected using stability ratios of 1 and 3, respectively. Each parameterization was calibrated to a melting temperature of ~343 K such that an effective temperature of 280(±10) K was employed for all folding simulations.

A unified-atom variant of the AMBER94 force field[36] was used to define bonded interactions (bond, angle, and torsion terms), with optimal geometries taken from the native state. To speed the dynamics of interest, low viscosity ($\tau_T = 0.5$ ps, ~2% that of water) was employed and dihedral energy terms were increased by a factor of 5 relative to the effective temperature. The polymer was made less rigid by also decreasing bonded interaction energies by a factor of 10 relative to the effective temperature. A complete set of simulation input files is available†.

All simulations were conducted using the stochastic dynamics integrator within the Gromacs molecular dynamics suite[37] with 15 Å cutoffs on all non-bonded interactions. As the mean fraction of native contacts present in simulations of the native state was 0.87, all folding simulations (started from a fully extended conformation) were considered fully folded when 87% of all possible secondary contacts and 87% of all possible tertiary contacts were formed.

Caveats of the biasing potential

While minimalist models allow the study of events not observable *via* fully atomistic molecular dynamics,[15,25,27] the limitations of the minimalist model approach should be considered. First of all, a quantitative prediction of absolute transition rates is not possible, and temporal analyses have therefore been conducted in units of iterations, yielding qualitative information on the nature of relative rates.[15] Moreover, since non-native interactions are less energetically favorable than native ones, this model is not expected to accurately characterize off-pathway folding intermediates (kinetic traps), other than those resulting from topological frustration.[15,26] However, we note that the inclusion of weak non-native attractions not present in previous biasing models greatly enhances the stochastic nature of folding trajectories, allowing a more plausible description of the underlying energetic landscape.

Furthermore, as ε_2 and ε_3 were chosen to be constant over the whole molecule, the simulated tRNA sequence may not have specific energetic biases to particular substructures. While some tRNA sequences may break this approximation, the goal of this work is to study the biasing role of polymeric entropy and we are thus investigating a general property of RNA folding, i.e. the effect of the tRNA topology on folding. While one could include substructure-dependent energetic biases to model differences between specific tRNA sequences, this would obscure the roles of polymeric entropy and topology in folding.

Finally, a prime consideration in employing a biasing potential of any sort, is the possibility of artifacts within the potential distorting the observed dynamics far from that of the true biomolecular system, and the true test of such a potential is the ability to show agreement with current knowledge prior to making predictions. Within the results reported above, we propose that such "artifactual folding" is seen as the early formation of the T_{12} tertiary contact (Figure 3(b)), which has not been shown to occur in the early stages of folding. Such potential artifacts may also affect the relative rates of individual helix-stem formation. We believe such artifactual folding to be minimal within the model, and stress that the ordering of

† http://folding.stanford.edu/tRNA

phase 1 secondary structure formation does not change the overall character of the analyses presented above.

Acknowledgements

We thank Erik Lindahl for technical aid, Dan Herschlag, Rhiju Das, Sung-Joo Lee, and Mark Engelhardt for their comments on the manuscript, and the Folding@Home volunteers worldwide who made this work possible (a list of contributors is available)†. G.J. is a Siebel Fellow. E.J.S. and Y.M.R. are supported by predoctoral fellowships from the Krell/DOE CSGF and Stanford Graduate Fellowship, respectively, with additional supporting grants ACS-PRF (36028-AC4) and NSF MRSEC CPIMA (DMR-9808677), and a gift from Intel.

References

1. Baker, D. (2000). A surprising simplicity to protein folding. *Nature*, **405**, 39–42.
2. Debe, D. A. & Goddard, W. A. (1999). First principles prediction of protein folding rates. *J. Mol. Biol.* **294**, 619–625.
3. Plaxco, K. W., Simons, K. T., Ruczinski, I. & David, B. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, **39**, 11177–11183.
4. Makarov, D. E., Keller, C. A., Plaxco, K. W. & Metiu, H. (2002). How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl Acad. Sci. USA*, **99**, 3535–3539.
5. Jewett, A. I., Pande, V. S. & Plaxco, K. W. (2003). Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J. Mol. Biol.* **326**, 247–253.
6. Klimov, D. K. & Thirumalai, D. (2000). Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl Acad. Sci. USA*, **97**, 7254–7259.
7. Gsponer, J. & Caflisch, A. (2001). Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.* **309**, 285–298.
8. Ferrara, P. & Caflisch, A. (2001). Native topology or specific interactions: what is more important for protein folding? *J. Mol. Biol.* **306**, 837–850.
9. Sorin, E. J., Rhee, Y. M., Nakatani, B. J. & Pande, V. S. (2003). Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophys. J.* **85**, 790–803.
10. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009.
11. Martínez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010–1016.
12. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016–1024.
13. Deschenes, L. A. & Vanden Bout, D. A. (2001). Single-molecule studies of heterogeneous dynamics in polymer melts near the glass transition. *Science*, **292**, 255–258.
14. McKinney, S. A., Declais, A. C., Lilley, D. M. J. & Ha, T. (2003). Structural dynamics of individual Holliday junctions. *Nature Struct. Biol.* **10**, 93–97.
15. Shimada, J. & Shakhnovich, E. I. (2002). The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl Acad. Sci. USA*, **99**, 11175–11180.
16. Zagrovic, B., Sorin, E. J. & Pande, V. (2001). β-Hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151–169.
17. Pande, V. S., Baker, I., Chapman, J., Elmer, S., Kaliq, S., Larson, S. *et al.* (2003). Atomistic protein folding simulations on the submillisecond timescale using worldwide distributed computing. *Biopolymers*, **68**, 91–109.
18. Sorin, E. J., Engelhardt, M. A., Herschlag, D. & Pande, V. S. (2002). RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* **317**, 493–506.
19. Thirumalai, D., Lee, N., Woodson, S. A. & Klimov, D. K. (2001). Early events in RNA folding. *Annu. Rev. Phys. Chem.* **52**, 751–762.
20. Shelton, V. M., Sosnick, T. R. & Pan, T. (2001). Altering the intermediate in the equilibrium folding of unmodified yeast $tRNA^{Phe}$ with monovalent and divalent cations. *Biochemistry*, **40**, 3629–3638.
21. Stein, A. & Crothers, D. M. (1976). Conformational changes of transfer RNA. The role of magnesium(II). *Biochemistry*, **15**, 160–168.
22. Stein, A. & Crothers, D. M. (1976). Equilibrium binding of magnesium(II) by *Escherichia coli* $tRNAf^{Met}$. *Biochemistry*, **15**, 157–160.
23. Ueda, Y., Taketomi, H. & Go, N. (1975). Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **7**, 445–459.
24. Ueda, Y., Taketomi, H. & Go, N. (1978). Studies on protein folding, unfolding, and fluctuations by computer-simulation. 2. 3-dimensional lattice model of lysozyme. *Biopolymers*, **17**, 1531–1548.
25. Shea, J. E., Onuchic, J. N. & Brooks, C. L. (2002). Probing the folding free energy landscape of the src-SH3 protein domain. *Proc. Natl Acad. Sci. USA*, **99**, 16064–16068.
26. Clementi, C., Hugh Nymeyer, H. & Onuchic, J. N. (2000). 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. *J. Mol. Biol.* **298**, 937–953.
27. Pande, V. S. & Rokhsar, D. S. (1998). Is the molten globule a third phase of proteins? *Proc. Natl Acad. Sci. USA*, **95**, 1490–1494.
28. Clementi, C., Garcia, A. E. & Onuchic, J. N. (2003). Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J. Mol. Biol.* **326**, 933–954.
29. Fang, X., Littrell, K., Yang, X.-J., Henderson, S. J.,

† Folding@Home, http://folding.stanford.edu

Siefert, S., Thiyagarajan, P. *et al.* (2000). Mg^{2+}-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*, **39**, 11107–11113.
30. Zhang, W. B. & Chen, S. J. (2002). RNA hairpin-folding kinetics. *Proc. Natl Acad. Sci. USA*, **99**, 1931–1936.
31. Ansari, A., Kuznetsov, S. V. & Shen, Y. Q. (2001). Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl Acad. Sci. USA*, **98**, 7771–7776.
32. Pande, V. S., Grosberg, A. Y. & Tanaka, T. (2000). Heteropolymer freezing and design: towards physical models of protein folding. *Rev. Mod. Phys.* **72**, 259–314.
33. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.* **4**, 497–502.
34. Nymeyer, H., Garcia, A. E. & Onuchic, J. N. (1998). Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl Acad. Sci. USA*, **95**, 5921–5928.
35. Li, L. & Shakhnovich, E. I. (2001). Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. *Proc. Natl Acad. Sci. USA*, **98**, 13014–13018.
36. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M. *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
37. Lindahl, E., Hess, B. & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317.

Edited by J. Doudna

(Received 3 October 2003; received in revised form 7 February 2004; accepted 10 February 2004)