Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures

Samuel S. Cho, David L. Pincus, and D. Thirumalai[1]

Biophysics Program, Institute for Physical Science and Technology, and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

Edited by Ken A. Dill, University of California, San Francisco, CA, and approved August 18, 2009 (received for review June 15, 2009)

Understanding how RNA molecules navigate their rugged folding landscapes holds the key to describing their roles in a variety of cellular functions. To dissect RNA folding at the molecular level, we performed simulations of three pseudoknots (MMTV and SRV-1 from viral genomes and the hTR pseudoknot from human telomerase) using coarse-grained models. The melting temperatures from the specific heat profiles are in good agreement with the available experimental data for MMTV and hTR. The equilibrium free energy profiles, which predict the structural transitions that occur at each melting temperature, are used to propose that the relative stabilities of the isolated helices control their folding mechanisms. Kinetic simulations, which corroborate the inferences drawn from the free energy profiles, show that MMTV folds by a hierarchical mechanism with parallel paths, i.e., formation of one of the helices nucleates the assembly of the rest of the structure. The SRV-1 pseudoknot, which folds in a highly cooperative manner, assembles in a single step in which the preformed helices coalesce nearly simultaneously to form the tertiary structure. Folding occurs by multiple pathways in the hTR pseudoknot, the isolated structural elements of which have similar stabilities. In one of the paths, tertiary interactions are established before the formation of the secondary structures. Our work shows that there are significant sequence-dependent variations in the folding landscapes of RNA molecules with similar fold. We also establish that assembly mechanisms can be predicted using the stabilities of the isolated secondary structures.

kinetic partitioning mechanism | parallel pathways | ribosomal frameshifting | RNA folding

The RNA folding problem has taken center stage in molecular biology because these molecules play a vital role in a variety of cellular functions. The percentage of the transcribed noncoding sequences in mice and human genomes exceed 90% (1), and the functional importance of the rest of the noncoding RNA is now only beginning to be understood (2). The noncoding roles of ribosomal RNA (rRNA) and transfer RNA (tRNA) are well known, but the discovery of ribozymes (RNA enzymes) has sparked an intense search for the functional roles of the rest of the noncoding RNA molecules (2, 3), which include catalysis, replication, transcriptional and translational regulation, and ligand binding (4), just to name a few. Consequently, it is important to determine the mechanisms by which RNA molecules fold so that a deeper understanding of the structure-function relationship can emerge.

Although great progress has been made in understanding of how RNA molecules fold (5−10) global principles that determine the sequence dependent folding mechanisms have not fully emerged. It is argued that RNA folding mechanisms must be inherently simple because of the apparent separation in the energy scales that describe the different levels of structural organization (11). Furthermore, RNA molecules are constructed from only four nucleotides (nts), and the Watson-Crick (WC) base pairings to some extent restrict the number of plausible secondary structures. Because the RNA helices are independently stable, RNA folding to a first approximation, can be described as being hierarchical in that the formation of the secondary structure precedes the acquisition of the tertiary structure. Typically the final tertiary structure does not significantly alter the preformed secondary structure (12). However, the chemical similarity of the nucleotides and the observation that only ≈50% of the nts are engaged in WC base pairing (13) results in the possibility that RNA can be trapped in a variety of alternate states even at the secondary structure level (14−16). The small stability gap separating the native basin of attraction (NBA) and the competing basins of attraction (CBAs) results in RNA being kinetically trapped in the CBAs during the folding process (5). Thus, even the folding landscape of RNA hairpins can be rugged, resulting in complex folding kinetics (17). In contrast to the diversity in the shapes and chemical characters of the building blocks of proteins, the four nts are all aromatic bases with similar chemical properties, thus making the interactions highly promiscuous. The low specificity of the interactions results in populations of alternate structures during the folding process (18–21).

To understand the inherent sequence-dependent variations in RNA folding, we performed coarse-grained simulations of three pseudoknots, which are involved in ribosomal frameshifting (22–25) and telomerase activity (26). Simulations using the Three Interaction Site (TIS) model for RNA (27) compare well with the available experimentally determined thermodynamic transitions. The folding landscapes and kinetics reveal the hidden complexity of RNA folding that manifests itself in the form of parallel folding routes. We explain the diversity of folding routes and the structural organization that occurs as the temperature changes in terms of sequence-dependent stabilities of the isolated helices, which we refer to as the "stability hypothesis."

Results and Discussion

Structure and Contact Maps for Pseudoknots. In the tertiary structure of a pseudoknot, nucleotides within a single stranded loop of a hairpin pair with nucleotides outside the loop, resulting in two double-stranded helical stems connected by single-standed loops. Equilibrium melting curves of pseudoknots involve one (28) or more transition temperature(s) (28–30), depending on their native structure, nucleotide sequence, and the valence of the counterions. We chose three pseudoknots that all have the most commonly occurring "H-type" (hairpin loop) motif (23). The MMTV and SRV-1 from viral genomes have a similar length with nearly identical contact maps [see supporting information (SI) text] that contain loops with no appreciable long-range (i.e., nucleotides that are well separated along the sequence) interactions (Fig. 1*A*–*F*). In

Author contributions: S.S.C. and D.T. designed research; S.S.C. and D.L.P. performed research; D.T. contributed new reagents/analytic tools; S.S.C. and D.T. analyzed data; and S.S.C. and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: thirum@umd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906625106/DCSupplemental.

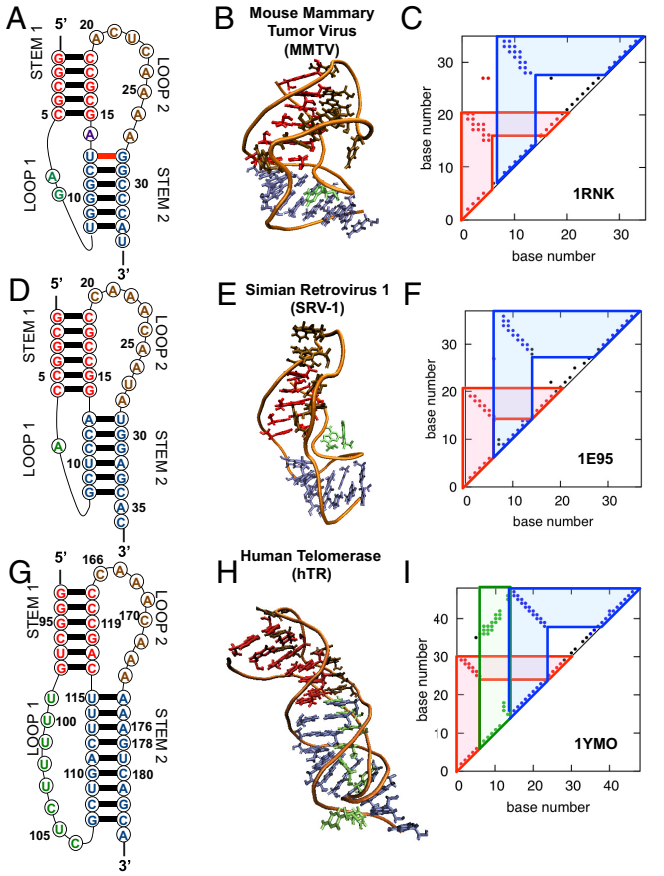

Fig. 1. Structural characteristics of the RNA pseudoknots. The secondary structure (left column), tertiary structures (center column), and contact maps (right column) are shown for RNA pseudoknots with PDB codes (*A–C*) 1RNK, (*D–F*) 1E95, and (*G–I*) 1YMO. The contact maps for the MMTV and SRV-1 pseudoknots are similar.

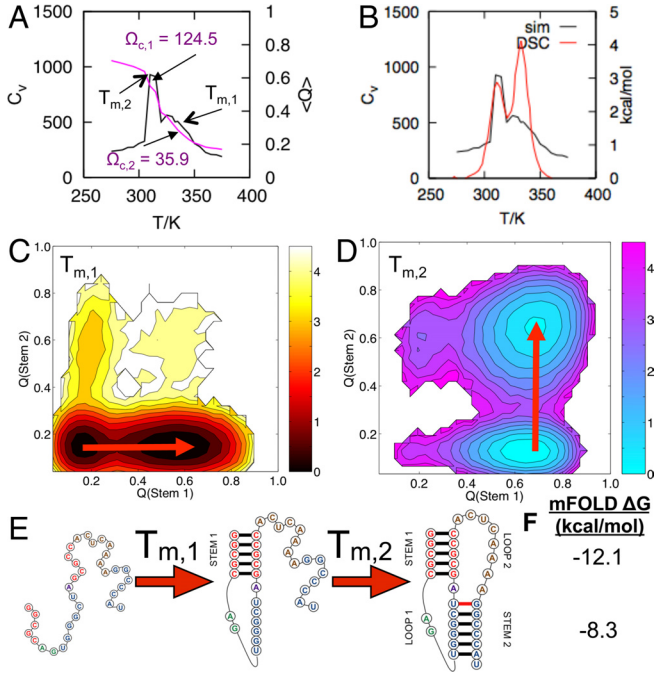

Fig. 2. Specific heat and free energy profiles for the MMTV pseudoknot. (*A*) Specific heat (C_v) versus temperature show two folding transition temperatures, $T_{m,1}$ and $T_{m,2}$. Also shown is the temperature dependence of $\langle Q \rangle$ (purple curve) and the corresponding cooperativity measure, $\Omega_{c,i}$, for the two transitions. (*B*) The C_v versus temperature profile from simulations is overlaid with the differential scanning calorimetry profile from Theimer and Giedroc (31) with the experimental temperature shifted down by 35 K. (*C* and *D*) The free energy profile $F(Q_1,Q_2)$ is projected onto the fraction of native contacts of the two stems (Q_1,Q_2) at (*C*) the higher folding transition temperature ($T_{m,1}$) and (*D*) the lower folding transition temperature ($T_{m,2}$). (*E*) Schematic diagram shows the structural transitions along the folding pathways inferred from $F(Q_1,Q_2)$. (*F*) The specific heat is measured in units of kcal/mol/K; free energies are in units of k_BT.

the case of the hTR, extracted from the human telomerase, one of the loops (Loop 1) makes a significant number of long-range tertiary interactions with Stem 2 (Fig. 1 *G* and *I*) that could play a significant role in determining folding cooperativity and pathways.

Thermal Transitions in MMTV and SRV-1. To disrupt the G-rich tracks in the MMTV sequence, Shen and Tinoco flipped four G-C base pairs (one from Stem 1 and three from Stem 2) to C-G base pairs (32) and designated the mutant as VPK. Because all available experimental characterization to date is for the VPK mutant, we will henceforth refer to the mutant as MMTV. Our simulations of the MMTV show two distinct melting temperatures in the C_v profile (Fig. 2*A*), in qualitative agreement with previous differential scanning calorimetry (DSC) profile (31) (Fig. 2*B*). The higher temperature, $T_{m,1}$, corresponds to the folding of the Stem 1 (Fig. 2*C*), whereas Stem 2 formation occurs at the lower temperature, $T_{m,2}$ (Fig. 2*D*). The sharpness of the second transition at $T_{m,2}$ relative to that at $T_{m,1}$ is linked to the simultaneous formation of Stem 2 and consolidates the tertiary interactions (see below). Insights into the structural transitions are obtained from the two-dimensional, temperature-dependent free energy landscape, $F(Q_1, Q_2)$, where Q_1 (Q_2) is the fraction of native contacts (see Materials and Methods for definition) associated with Stem 1 (Stem 2). It is clear from Fig. 2*C* that at $T = T_{m,1} \approx 325$ K the hairpin associated with Stem 1 forms, and the folding of the second stem occurs at $T = T_{m,2} \approx 305$ K (Fig. 2*D*). The separation in the two melting temperatures by about 20 K shows that the thermal folding of the MMTV is effectively hierarchical (Fig. 2*E*). The two helices form independently and coalesce to form the pseudoknot. However, within this hierarchical picture, there are parallel pathways as the explicit kinetic simulations show (see below).

The simulations can be compared with melting curves obtained using ultraviolet absorbance and by differential scanning calorimetry (29, 31). The values of the two melting temperatures for the MMTV at high NaCl concentration, which is most appropriate for the model used here because the stacking interaction energies were obtained at 1M NaCl, are $T_{m,2} = 347$ K and $T_{m,1} = 368$ K, with the $\Delta T_m = T_{m,1}$-$T_{m,2} = 21$ K (31) (Fig. 2*B*). Given the simplicity of the TIS model and the plausible sequence effects, we surmise that the simulated results for the C_v are in reasonable agreement with experiments. The difference between $T_{m,1}$ and $T_{m,2}$ ($\Delta T_m \approx 21$ K) found in experiments is in quantitative agreement with the simulation results ($\Delta T_m \approx 20$ K).

In contrast to the MMTV, there is only one peak in C_v (Fig. 3*A*) for the SRV-1. Given the similarity in the contact maps of MMTV and SRV-1 (compare Fig. 1 *C* and *F*), the cooperative ordering of Stems 1 and 2 and the formation of the pseudoknot in SRV-1 compared with the hierarchical assembly in MMTV is surprising. The number of GC pairs in Stem 2 of MMTV and SRV-1 is the same (Fig. 1 *A* and *D*). However, there is a GU mispairing in Stem 2 of MMTV, whereas the nts in Stem 2 of SRV-1 follow the WC pairing. Thus, we expect that the isolated helix corresponding to Stem 2 in MMTV could be less stable than helix 2 from SRV-1 (see below), which would in turn compromise the cooperativity of MMTV.

To quantify the cooperativity of the equilibrium transitions, we

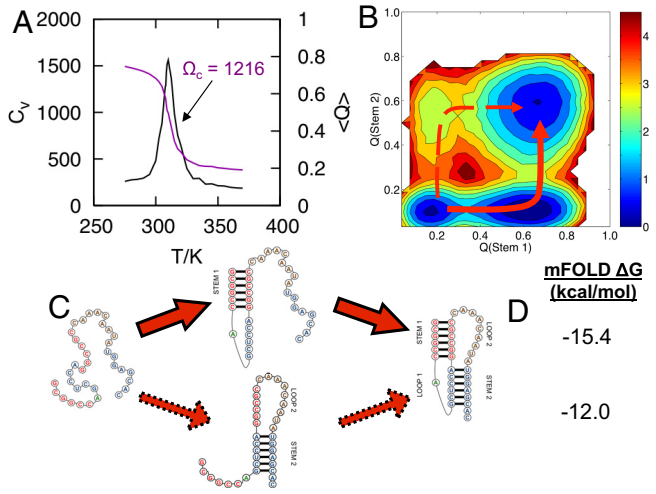

Fig. 3. Thermodynamics characteristics for the SRV-1 pseudoknot. (*A*) Specific heat (C_v) versus temperature shows a single folding transition temperature. Dependence of $\langle Q \rangle$ on the temperature and the corresponding cooperativity measure, Ω_c, show that the SRV-1 formation is cooperative. (*B*) Free energy profile, $F(Q_1,Q_2)$, is projected onto the fraction of native contacts of the two stems (Q_1,Q_2). (*C*) The overall folding mechanism predicted by $F(Q_1,Q_2)$ shows that the dominant folding pathway involves the formation of Stem 1 followed by Stem 2. (*D*) Stabilities of the individual stems predicted by Mfold.

used a measure $\Omega_{c,i} = \frac{T_{m,i}^2}{\Delta T_{m,i}} \left| \frac{d\langle Q \rangle}{dT} \right|_{m,i}$, where $\left| \frac{d\langle Q \rangle}{dT} \right|_{m,i}$ is evaluated at $T = T_{m,i}$ and $\Delta T_{m,i}$ is the width at the half-maximum of $\left| \frac{d\langle Q \rangle}{dT} \right|_{m,i}$ at $T = T_{m,i}$ (33). For MMTV, $\Omega_{c,2} \approx 36$ and $\Omega_{c,1} \approx 125$, whereas at the single melting transition of SRV-1 $\Omega_c \approx 1216$. The large value of Ω_c for SRV-1 suggests all of the interactions that stabilize the folded structure form cooperatively, which obscures the hidden structural transitions in $C_v(T)$. However, the details of the ordering process can be discovered from the folding landscape, $F(Q_1, Q_2)$ (Fig. 3*B*), which shows that the formation of Stem 2 populates a high free energy intermediate. Thus, the most probable way in which SRV-1 forms is by initial formation of Stem 1, which nucleates the rest of the structure over a narrow temperature range that may be hard to delineate from specific heat alone. The equilibrium folding studies of the SRV-1 have not been performed experimentally, so our results constitute a readily verifiable prediction.

Structural Transitions in hTR. The biological activities of telomerase, a ribonucleoprotein complex that contains a pseudoknot, include the upregulation of a vast majority of cancer lines (34) and a link between aging and telomere length (35). The hTR pseudoknot has a short GC-rich stem (Stem 1) and a longer AU-rich stem (Stem 2), and one of the loops (Loop 1) forms a significant number of long-range pseudoknot interactions with Stem 2 (Fig. 1 *G–I*). The uridine-rich Loop 1 is universally conserved among vertebrates, and it consists of extensive Hoogsteen base triples with Stem 2 that form a short series of pyrimidine-purine-pyrimidine RNA triplex structures (28).

The wild-type hTR, which contains the U177 single-nucleotide bulge (Fig. 1*g*), is in dynamic equilibration with a hairpin conformation found between nts 93–121 (36). Deletion of U177 stabilizes the pseudoknot, thus eliminating the switch to the alternate conformation (28). We simulated the ΔU177 construct, which we refer to as hTR (Fig. 1 *G–I*). The temperature dependence C_v shows a single sharp peak at $T_m \approx 315$ K, which shows that the formation of the hairpins and the formation of the pseudoknot (tertiary) interactions are tightly coupled. However, examination of the free

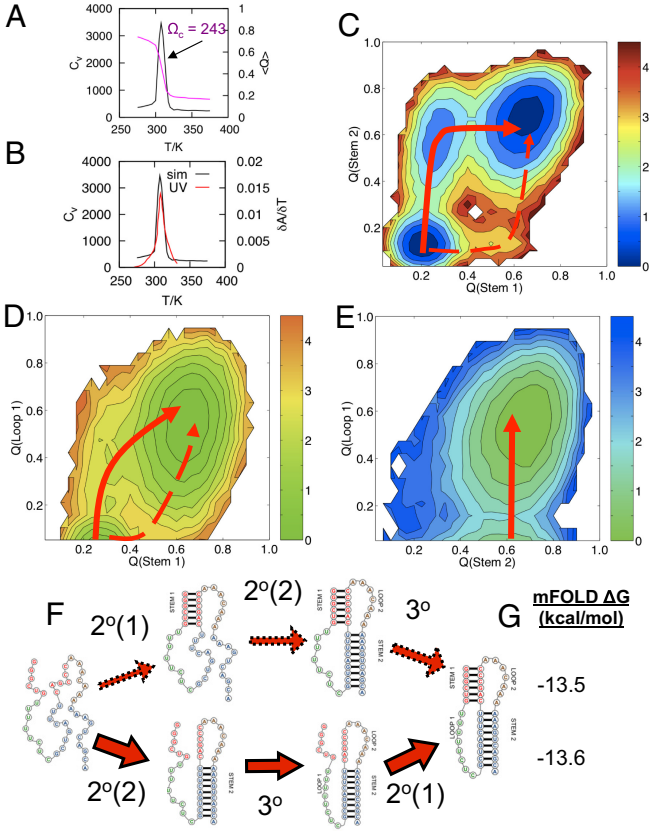

Fig. 4. Equilibrium transitions in the hTR pseudoknot. (*A*) Same as Fig. 2*A* except that there is only one peak in the temperature dependence of $C_v(T)$. The $\langle Q \rangle$ as a function of T and the associated Ω_c value are given. (*B*) Same as Fig. 2*B* except that the ultraviolet melting profile from (28) is shown. (*C*) Same as Fig. 2*C*. (*D*) and (*E*) correspond to the free energy profiles in terms of order parameters (Q_{L1},Q_1) and (Q_{L1},Q_2), respectively. Here, Q_{L1} is the number of tertiary contacts that Loop 1 (Fig. 1*G*) makes with the rest of the pseudoknot. (*F*) Set of folding pathways obtained from the free energy profiles are outlined. (*G*) Free energies of the constructs associated with the helices computed using Mfold are shown.

energy profiles, $F(Q_1, Q_2)$, indicates that structural ordering transitions in the hTR is complex. One possible way that hTR assembles is through the formation of Stem 1, although it is also likely that Stem 2 can form a hairpin before further consolidation of the structure (Fig. 4*C*). When Stem 1 forms first, the long-range tertiary interactions in Loop 1 is also structured even before the formation of Stem 2. In another competing scheme, Stem 2 is fully formed before development of structure in Loop 1. Interestingly, in this alternate way in which the structure assembles, a certain amount of tertiary interactions must be established before the complete formation of the secondary structure.

The structural changes that occur as hTR melts, which can be gleaned using the free energy profiles (Fig. 4 *C–E*), are not clearly reflected in the calorimetric specific heat from simulations nor the UV melting profiles from experiments (28) (Fig. 4 *A* and *B*). The structural transitions shown in Fig. 4 *A–E* are consistent with the inferences drawn from the UV melting curves for the ΔU177 construct (28). It was proposed, based on the UV melting curves, that there are three melting transitions that correlate with the loss of tertiary structure, followed by melting of the AU rich Stem 2, and eventually the loss of structure in Stem 1 (see preferred route in Fig. 4*F*). The peak in $C_v(T)$ occurs at $T \approx 42$ °C in Fig. 4*A*, which, by comparison with the first derivative of the UV melting profiles, can be associated with the tertiary structure formation. Experiments show a shoulder at $T \approx 45$ °C and a sharper peak at $T \approx 70$ °C (28),

which is not captured in our simulations. However, the hidden structural transitions can be inferred from $F(Q_1, Q_2)$ as described above. Thus, by combining the $C_v(T)$ profile and $F(Q_1, Q_2)$, we can identify the structural transitions in hTR that is qualitatively in accord with experiments.

Thermodynamic Stabilities of the Isolated Helices Determine the Order of the Structural Transitions. The folding landscapes (Figs. 2–4) lead to the stability hypothesis: namely, the order of formation of the helices, as well as the subsequent consolidation to form tertiary structures, is largely determined by the stabilities of the isolated helices. To make this observation precise, we determined the stabilities of the isolated helices in the pseudoknots. We constructed the isolated helices using the core that is connected by a GCAA loop (Fig. S1). The free energies of the helices determined using the Mfold algorithm, which is accurate in obtaining the free energies of the optimal and suboptimal structures of small hairpins, show that because of the presence of one additional GC pair, Stem 1 (Fig. 2*A*) in the MMTV is more stable than Stem 2 by about 3.8 kcal/mol (Fig. 2*F*). If the stability of the helices is the major determining factor in the assembly of pseudoknots, we expect that Stem 1 of MMTV should melt at the higher temperature than Stem 2. Such an inference is in accord with our simulations that clearly show that at $T_{m,1}$ ($> T_{m,2}$), the more stable helix corresponding to Stem 1 forms. Only at a lower temperature, $T_{m,2}$, does the less stable Stem 2 acquire its entire structure.

For the SRV-1, the Mfold algorithm predicts that Stem 1, a GC rich region, is more stable than Stem 2 by $\Delta G = -3.4$ kcal/mol. The relative stability, $\Delta\Delta G/\Delta G_1 = (\Delta G_2\text{-}\Delta G_1)/\Delta G_1$ ($\Delta G_1 = -15.4$ kcal/mol) is 0.2 compared with 0.3 for MMTV. The stability hypothesis suggests that Stem 1 should form at a higher temperature than Stem 2. As explained earlier, the large cooperativity associated with pseudoknot formation makes it difficult to draw this conclusion from $C_v(T)$ alone (Fig. 3*A*) but is clear from $F(Q_1, Q_2)$ (Fig. 3*b*).

Interestingly, the stabilities of Stem 1, Stem 2 + Loop1, and Stem 2 in isolation in hTR are all similar (Fig. S1), which implies that there should be multiple ways in which the pseudoknot can assemble (see below). From a thermodynamic perspective, we expect that the cooperativity of the formation should also be compromised. The value of Ω_c for hTR is about 243, which is not as large as found for SRV-1. Thus, the stabilities of the individual helices are reasonable indicators of how RNA pseudoknots assemble from thermodynamic and kinetic perspectives (see below).

Folding Kinetics: Parallel Pathways and Hierarchical Assembly. In the interpretation of the thermally induced structural transitions using the equilibrium free energy profiles (Figs. 2–4), we have alluded to the plausible connections to the routes explored by the pseudoknots with the features in $F(Q_1,Q_2)$. Such a link between equilibrium quantities and the folding kinetics was previously established in the folding of *Tetrahymena* ribozyme using experiments and theory, which showed that the extent of cooperativity and the kinetics of folding of RNA are anti-correlated (19). In particular, it was argued that as the thermodynamic cooperativity increases, the diversity of the routes decrease because all of the interactions that stabilize the native structure are consolidated in a single step (19). On the other hand, if the secondary structural elements have comparable stability, then there are multiple routes to the native state.

In general, we expect that folding in a rugged RNA landscape can be described in terms of the kinetic partitioning mechanism (KPM) (19), according to which a fraction of molecules, Φ, reaches the NBA directly, whereas the remaining fractions (1-Φ) becomes trapped in CBAs. In the case of pseudoknots, the free energy profiles, $F(Q_1, Q_2)$, suggest that the distinct parallel routes to the NBA can exist even when trapping in the CBAs is small because of the elimination of energetic frustration. Nevertheless, the equilibrium free energy profiles in Figs. 2–4 suggest that the complex folding through multiple channels to the NBA can be described using $P_u(t)\Sigma_i^{N_P}\Phi_i\exp\left(-\frac{t}{\tau_i}\right)$ where $P_u(t)$ is the fraction of molecules that remains unfolded at time, t, τ_i is the time constant for reaching the NBA via the set of pathways, i, Φ_i is the corresponding flux in i, and N_p is the number of discernible pathways. If the maximum flux through a particular channel $\Phi_{max} \approx 1$, then RNA folding occurs through a hierarchical mechanism, whereas $0 < \Phi_{max} < 1$ implies parallel routes to the NBA.

To corroborate the expectation of parallel folding routes to the NBA inferred from $F(Q_1,Q_2)$, we generated 100 folding trajectories for each of the pseudoknots by integrating the Langevin equations of motion. To model the overdamped motion, we used a friction coefficient that corresponds to the viscosity of water (see Methods and *SI Text* for details). Folding is triggered by initially preparing an ensemble of unfolded conformations at high temperature (T) and quenching T to a value less than the lowest melting temperature. For each trajectory, we calculated the first passage time to the native state using Q as the progress variable. From a distribution of the first passage times, $P_{FP}(s)$, for the 100 folding trajectories, we calculated $P_u(t) = 1 - \int_0^t P_{FP}(\mathrm{S})ds$.

MMTV. In accordance with the predictions based on $F(Q_1,Q_2)$, MMTV folds in a hierarchical manner with parallel routes (Fig. 5 *B* and *C*). The time dependence of $P_u(t)$ is fit using a sum of two exponentials (Fig. 5). The flux of molecules through the dominant pathway, in which Stem 1 folds before Stem 2, is $\Phi_{12} = 0.77$, whereas the remaining fraction ($\Phi_{21} = 0.23$) folds by the alternate route predicted using $F(Q_1,Q_2)$. The time-dependent structural transitions that occur as MMTV folds are also observed (Fig. 5 *B* and *E*). If the stability of the isolated helices also determines the kinetics of folding, we expect that Stem 1 should form more rapidly than Stem 2. Indeed the simulations show that helix 1 forms with a time constant around $\tau_{12} = 0.4$ ms (Fig. 5*A*), whereas the time constant for the folding of Stem 2 is $\tau_{21} = 4.3$ ms. A similar scenario occurs in the folding of the larger *Tetrahymena* ribozyme in which the more stable P4-P6 subdomain folds before the formation of the rest of the structure (5, 8). We conclude that the ordering of Stem 1 nucleates the formation of the tertiary structure. The preformation of Stem 2, which naturally occurs in the second channel, does not affect the rate of formation of Stem 1, as shown by the analysis of the 23 trajectories that fold by the alternate route (Fig. S2 *a* and *b*). The independent formation of the two helices shows that MMTV folds by parallel routes.

SRV-1. In contrast to MMTV, the structurally similar SRV-1 (compare Fig. 1 *C* and *F*) folds in a single step when analyzed in terms of $P_u(t)$ (Fig. 5 *E* and *F*). From the double exponential fit of $P_u(t)$ (Fig. 5*D*), we find that only 6% of the molecules reached the NBA by the less probable route (i.e., Stem 2 forms before Stem 1) (Fig. 5*F*). Analyses of individual trajectories show that in the dominant route ($\Phi_{12} = 0.94$) the more stable Stem 1 folds, which subsequently drives the formation of Stem 2 and the tertiary structure (Fig. 5*E*). Within the dominant channel, the time for helix formation in the GC-rich Stem 1 is comparable to Stem 1 in MMTV, which establishes that helix formation largely depends only on the local sequence (Fig. S3 *a* and *c*). The small probability of observing alternative pathways to the NBA is a consequence of the high cooperativity in the tertiary structure formation ($\Omega_c \approx 1216$ for SRV-1, whereas only $\Omega_{c,1} \approx 125$ and $\Omega_{c,2} \approx 36$ for MMTV. In addition, $F(Q_1,Q_2)$ for SRV-1 shows that the route corresponding to the formation of Stem 2 before Stem 1 has a higher free energy barrier, which makes it kinetically improbable. Thus, the explicit kinetic simulations are in agreement with predictions made using $F(Q_1,Q_2)$.

hTR. The complex assembly mechanism inferred using the free energy profiles (Fig. 4 *B* and *D*) are reflected, to a large extent, in the folding kinetics of the hTR. From the stabilities of the second-

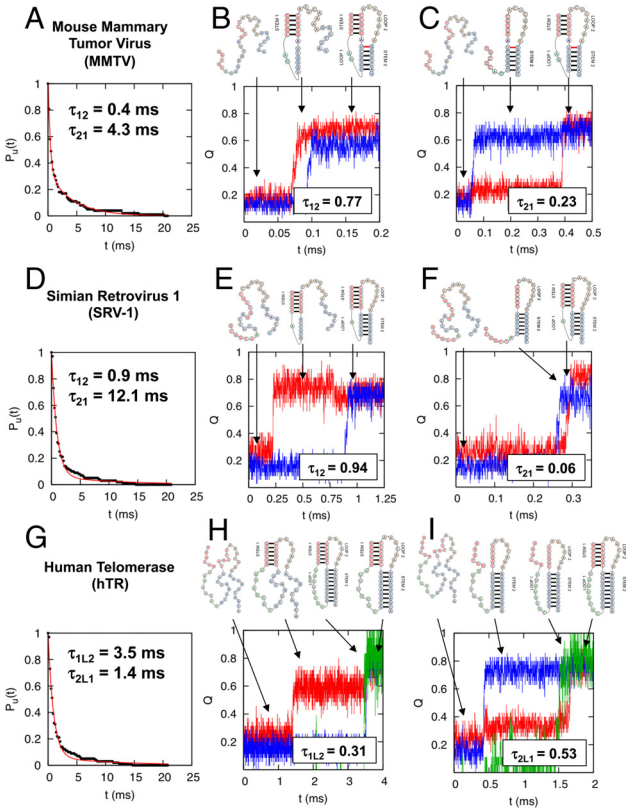

Fig. 5. Folding kinetics of the pseudoknots. (*A*) Dependence of the fraction of unfolded molecules, $P_u(t)$, as a function of time for the MMTV pseudoknot. Solid red line is a double exponential fit of $P_u(t)$ with time constant τ_{12} (Stem 1 forms before Stem 2) and τ_{21} (Stem 2 forms before Stem 1). (*B*) Typical trajectory for the dominant pathway (amplitude $\Phi_{12} = 0.77$) that shows the acquisition of the total number of contacts, *Q*. Stem 1 (red) forms before Stem 2 (blue). The structures that form at different stages are shown above the trajectory. (*C*) In the subdominant pathway (amplitude $\Phi_{12} = 0.23$), Stem 2 (blue) forms before Stem 1 (red). Structures associated with the pathway are shown above the trajectory. (*D–F*) Same as (*A–C*) except that these graphs describe the folding kinetics of the SRV-1 pseudoknot. (*G*) Same as (*A*) except that the double-exponential fit to $P_u(t)$ is for the two most dominant pathways for the hTR pseudoknot. Both (*H*) and (*I*) show the time dependence of Q for the two dominant pathways with the amplitudes Φ_{1L2} (the pathway with the order of formation of the pseudoknot is Stem 1, tertiary interactions associated with L1, and then the ordering of Stem 2) and Φ_{2L1} (the order of formation of the pseudoknot is Stem 2, L1, and then Stem 1). The structures associated with these transitions are displayed at the *top*. The remaining set of possible pathways for the hTR pseudoknot have the following values of Φ_i: $\Phi_{12L} = 0.02$, $\Phi_{21L} = 0.07$, $\Phi_{L12} = 0.03$, and $\Phi_{L21} = 0.04$.

ary structures and the thermodynamic analyses, we proposed that the folding of hTR should involve parallel routes including the possibility that tertiary interactions can form before assembly of the helices. The time-dependence of $P_u(t)$ reflects the anticipated richness in the folding kinetics. There are two dominant channels (Fig. 5 *H* and *I*) with $\Phi_{2L1} = 0.53$ (i.e., Stem 2 folds, followed by interactions involving Loop 2, which is the lower route predicted in Fig. 4*E*) and $\Phi_{1L2} = 0.34$ (i.e., formation of Stem 1 precedes the ordering of Loop 2). The rest of the molecules (16%) followed one of four distinct channels (Fig. 5). In the channel with $\Phi_{2L1} = 0.53$, tertiary interactions are established before the formation of Stem 1 as shown by the time dependence of *Q* (Fig. 5*H*). Here, interactions of Loop 1 (green curve in Fig. 5*H*) are fully formed before the formation of Stem 2 (red curve in Fig. 5*H*). Thus, in this route, the standard scenario that secondary structure formation occurs before assembly of tertiary structures is not observed. In the second most dominant route, hTR forms by a hierarchical mechanism. We should point out in the kinetic assembly of hTR, the less stable helix (as inferred from thermodynamic considerations) forms preferentially by a factor of $\Phi_{2L1}/\Phi_{1L2} \approx 1.6$. Thus, the free energy profiles can give only a qualitative picture of the assembly mechanism, and hence detailed kinetic studies are needed to obtain a molecular picture of the folding process.

Discussion

To determine the principles that govern RNA self-assembly at the sequence level, we performed systematic thermodynamic and kinetic studies of coarse-grained molecular simulations. Although the coarse-grained (CG) models have a number of limitations, they have been proved useful in deciphering the key features that control the folding of a variety of RNA molecules (27, 37–40). Comparisons of our simulations of RNA pseudoknots to previously observed equilibrium optical and calorimetric melting profiles show excellent agreement. Our results show how variations in the sequence can greatly alter the folding mechanism of pseudoknots with very similar folded structures as judged by the contact maps, which is in sharp contrast to what is generally observed to be the case in protein folding. Dramatic variations in the kinetics of assembly are reflected in the sequence-dependent changes in the equilibrium free energy profiles, which point to complex folding mechanisms involving intermediates and parallel pathways to the NBA.

It is thought that typically RNA tertiary structure formation occurs by assembly of preformed secondary structure because the latter are independently stable (12). One of our key findings is that even when hierarchical assembly is observed, as is the case for the MMTV, there are parallel channels to the NBA that are reflected only when the kinetics of assembly is examined in detail. The existence of parallel routes to the NBA, even when trapping in metastable states is improbable, implies that there are multiple transition states. The folding mechanism of the highly cooperative folding of SRV-1 shows there is essentially a single pathway to the native state. Interestingly, folding of the SRV-1 shows that there is an inverse correlation between the extent of thermodynamic cooperativity and diversity in the folding pathways. In contrast, hTR folds with several parallel routes to the NBA. Such a finding is in accord with a theoretical study in which it was argued that if there is a balance between the interactions stabilizing the various structural elements (as is the case in the hTR), then folding occurs efficiently by multiple pathways (41). Folding of hTR further shows that tertiary contacts can form before the complete assembly of secondary structures (Fig. 5 *H* and *I*), so these loops can be significant in determining folding cooperativity and pathways in some cases, although secondary structure base pairing is expected to be dominant. Although the generality of the finding that tertiary interactions can consolidate secondary structure formation is not clear, there are examples that suggest that this scenario can be realized in several RNA molecules (15, 30, 42).

The relative stabilities of the constituent secondary structures (the helices that form stems in our case studies) are the main determinant of the ordering mechanisms of RNA. It has long been appreciated that the disparity in the stabilities of secondary and tertiary interactions allows one to view RNA folding in terms of hierarchical assembly of preformed secondary structures. However, in the present study, we systematically and unambiguously establish the quantitative relationship between the relative stabilities of helices and the kinetics of the folding process itself. If the stems are sufficiently different in stabilities with distinct melting temperatures, the more stable one folds first, resulting in a hierarchical folding mechanism with parallel pathways. If the stems are similar in stability with nearly identical melting temperatures, then a parallel folding mechanism results. Because the stabilities of isolated helices are determined by the sequence, the folding landscape itself can be dramatically altered by mutations. It also follows that detailed studies of the constituent helices can foretell how a complete RNA molecule with a well-defined tertiary structure can fold. It should

be stressed that there are multiple RNA folding scenarios that can be dissected in terms of the stabilities of individual helices. Indeed, even for structurally related pseudoknots, the assembly pathways can be distinct.

Materials and Methods

The energy function for the TIS model and the parameters are described in detail elsewhere (27). The specific heats were calculated using multiple trajectories in which a number of transitions between all of the relevant states were observed. Similar results were obtained using the weighted histogram analysis method (WHAM) (Fig. S3). To probe the folding kinetics, we generated 100 trajectories for each pseudoknot by integrating the Langevin equation using the value of friction that is appropriate for water. The simulation details including the calculation of free energies of the isolated using Mfold are given in the *SI Text*.

ACKNOWLEDGMENTS. The authors are grateful for useful discussions with Prof. Changbong Hyeon. This work was supported in part by a grant from the National Science Foundation (CHE 09–14033). S.S.C. and D.L.P. each appreciate support from Ruth L. Kirschstein National Research Service Awards from the National Institutes of Health.

1. Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. *Eur J Hum Genet* 13:894–897.
2. Prasanth KV, Spector DL (2007) Eukaryotic regulatory RNAs: An answer to the 'genome complexity' conundrum. *Genes Dev* 21:11–42.
3. Doudna JA, Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* 418:222–228.
4. Storz G (2002) An expanding universe of noncoding RNAs. *Science* 296:1260–1263.
5. Treiber DK, Williamson JR (1999) Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* 9:339–345.
6. Sosnick TR, Pan T (2003) RNA folding: Models and perspectives. *Curr Opin Struct Biol* 13:309–316.
7. Thirumalai D, Hyeon C (2005) RNA and protein folding: Common themes and variations. *Biochem* 44:4957–4970.
8. Woodson SA (2005) Structure and assembly of group I introns. *Curr Opin Struct Biol* 15:324–330.
9. Laederach A, Shcherbakova I, Liang MP, Brenowitz M, Altman RB (2006) Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule. *J Mol Biol* 358:1179–1190.
10. Chen SJ (2008) RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* 37:197–214.
11. Tinoco I, Jr, Bustamante C (1999) How RNA folds. *J Mol Biol* 293:271–281.
12. Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137.
13. Dima RI, Hyeon C, Thirumalai D (2005) Extracting stacking interaction parameters for RNA from the data set of native structures. *J Mol Biol* 347:53–69.
14. Henley DD, Lindahl T, Fresco JR (1966) Hydrodynamic changes accompanying the thermal denaturation of transfer ribonucleic acid. *Proc Natl Acad Sci USA* 55:191–198.
15. Wu M, Tinoco I, Jr (1998) RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci USA* 95:11555–11560.
16. Thirumalai D (1998) Native secondary structure formation in RNA may be a slave to tertiary folding. *Proc Natl Acad Sci USA* 95:11506–11508.
17. Hyeon C, Thirumalai D (2008) Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins. *J Am Chem Soc* 130:1538–1539.
18. Thirumalai D, Woodson SA (1996) Kinetics of folding of proteins and RNA. *Acc Chem Res* 29:433–439.
19. Pan J, Thirumalai D, Woodson SA (1997) Folding of RNA involves parallel pathways. *J Mol Biol* 273:7–13.
20. Treiber DK, Rook MS, Zarrinkar PP, Williamson JR (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science* 279:1943–1946.
21. Chen SJ, Dill KA (2000) RNA folding energy landscapes. *Proc Natl Acad Sci USA* 97:646–651.
22. Brierley I (1995) Ribosomal frameshifting viral RNAs. *J Gen Virol* 76:1885–1892.
23. Giedroc DP, Theimer CA, Nixon PL (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* 298:167–185.
24. Harger JW, Meskauskas A, Dinman JD (2002) An ''integrated model'' of programmed ribosomal frameshifting. *Trends Biochem Sci* 27:448–454.
25. Cao S, Chen SJ (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* 34:2634–2652.
26. Theimer CA, Feigon J (2006) Structure and function of telomerase RNA. *Curr Opin Struct Biol* 16:307–318.
27. Hyeon C, Thirumalai D (2005) Mechanical unfolding of RNA hairpins. *Proc Natl Acad Sci USA* 102:6789–6794.
28. Theimer CA, Blois CA, Feigon J (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* 17:671–682.
29. Gonzalez RL, Jr, Tinoco I, Jr (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J Mol Biol* 289:1267–1282.
30. Gluick TC, Wills NM, Gesteland RF, Draper DE (1997) Folding of an mRNA pseudoknot required for stop codon readthrough: Effects of mono- and divalent ions on stability. *Biochem* 36:16173–16186.
31. Theimer CA, Giedroc DP (2000) Contribution of the intercalated adenosine at the helical junction to the stability of the gag-pro frameshifting pseudoknot from mouse mammary tumor virus. *RNA* 6:409–421.
32. Shen LX, Tinoco I, Jr (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J Mol Biol* 247:963–978.
33. Klimov DK, Thirumalai D (1998) Cooperativity in protein folding: From lattice models with sidechains to real proteins. *Fold Des* 3:127–139.
34. Blasco MA (2003) Telomeres and cancer: A tale with many endings. *Curr Opin Gen Dev* 13:70–76.
35. Marciniak RA, Johnson FB, Guarente L (2000) Dyskeratosis congenita, telomeres and human ageing. *Trends Genet* 16:193–195.
36. Theimer CA, Finger LD, Trantirek L, Feigon J (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc Natl Acad Sci USA* 100:449–454.
37. Sorin EJ, et al. (2004) Does native state topology determine the RNA folding mechanism? *J Mol Biol* 337:789–797.
38. Pincus DL, Cho SS, Hyeon C, Thirumalai D (2008) Minimal models for proteins and RNA from folding to function. *Prog Mol Biol Trans Sci* 84:203–250.
39. Ding F, et al. (2008) Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* 14:1164–1173.
40. Lin JC, Thirumalai D (2008) Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *J Am Chem Soc* 130:14080–14081.
41. Thirumalai D, Woodson SA (2000) Maximizing RNA folding rates: A balancing act. *RNA* 6:790–794.
42. Chauhan S, Woodson SA (2008) Tertiary interactions determine the accuracy of RNA folding. *J Am Chem Soc* 130:1296–1303.