doi:10.1016/j.jmb.2007.01.006 J. Mol. Biol. (2007) 367, 909–924

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Biphasic Folding Kinetics of RNA Pseudoknots and Telomerase RNA Activity

Song Cao and Shi-Jie Chen*

[1]*Department of Physics University of Missouri-Columbia, Columbia MO 65211, USA*

[2]*Department of Biochemistry University of Missouri-Columbia, Columbia MO 65211, USA*

**Corresponding author*

Using a combined master equation and kinetic cluster approach, we investigate RNA pseudoknot folding and unfolding kinetics. The energetic parameters are computed from a recently developed *Vfold* model for RNA secondary structure and pseudoknot folding thermodynamics. The folding kinetics theory is based on the complete conformational ensemble, including all the native-like and non-native states. The predicted folding and unfolding pathways, activation barriers, Arrhenius plots, and rate-limiting steps lead to several findings. First, for the PK5 pseudoknot, a misfolded 5′ hairpin emerges as a stable kinetic trap in the folding process, and the detrapping from this misfolded state is the rate-limiting step for the overall folding process. The calculated rate constant and activation barrier agree well with the experimental data. Second, as an application of the model, we investigate the kinetic folding pathways for human telomerase RNA (hTR) pseudoknot. The predicted folding and unfolding pathways not only support the proposed role of conformational switch between hairpin and pseudoknot in hTR activity, but also reveal molecular mechanism for the conformational switch. Furthermore, for an experimentally studied hTR mutation, whose hairpin intermediate is destabilized, the model predicts a long-lived transient hairpin structure, and the switch between the transient hairpin intermediate and the native pseudoknot may be responsible for the observed hTR activity. Such finding would help resolve the apparent contradiction between the observed hTR activity and the absence of a stable hairpin.

Keywords: kinetics; RNA pseudoknot; activation energy; misfolded state; telomerase

Introduction

An RNA pseudoknot structure is formed through base-pairing between loop and other regions of RNA such as loops or unpaired 5′ or 3′ ends. The simplest RNA pseudoknot is the H (hairpin)-type pseudoknot. As shown in Figure 1(a) and (b), an H-pseudoknot consists of two stems, S_1 and S_2, and two loops, L_1 and L_2, that span across the deep (major) groove of S_2 and the shallow (minor) groove of S_1, respectively. RNA pseudoknots play critical roles in a wide range of biological functions, from ribosomal frameshifting[1–7] to human telomerase RNA (hTR) activity[8–15] and hepatitis delta virus ribozyme functions.[16–18] Several experiments have clearly demonstrated the essential kinetic roles in pseudoknot functions. For example, in ribosomal pausing at messenger RNA slippery sites and the −1 frameshifting,[4] where pseudoknot plays a central role, the time scale of pseudoknot formation/disruption is one of the key factors to determine the frameshifting efficiency.[1,3] In hTR functions, genetic and biochemical experiments revealed the critical functional role of a molecular switch between hairpin and pseudoknot of hTR.[9,10] For instance, mutants that alter the hairpin-pseudoknot equilibrium would cause hTR to lose activity, resulting in disease such as dyskeratosis congenita.[9–11,14,15] On the other hand, in a separate mutational study, a hairpin-destabilizing mutation was found to retain hTR activity, raising questions on the functional role of hairpin-pseudoknot switch.[13] It would be intriguing to understand whether such a mutation can

Abbreviation used: hTR, human telomerase RNA.

E-mail address of the corresponding author: chenshi@missouri.edu

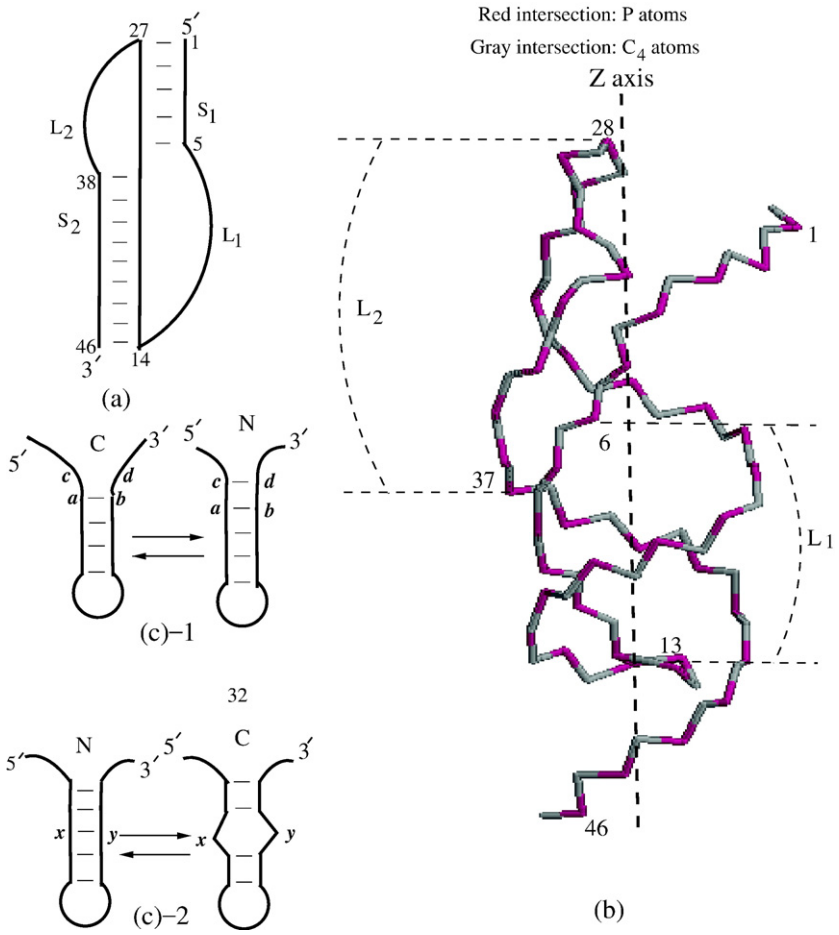

Figure 1. A simplified secondary structure representation (a) and C_4 and P virtual bond representations[32,62] (b) for a pseudoknot consist of two helix stems (S_1 and S_2) and two loops (L_1 and L_2). (c) Kinetic moves between two kinetically adjacent conformation states are defined as the formation and breaking of a base stack (c)-1 or a stacked base-pair (c)-2.

possibly result in a kinetically (instead of thermodynamically) populated hairpin intermediate state. The ability to quantitatively predict pseudoknot folding kinetics from the nucleotide sequence would be essential for the deconvolution of the molecular mechanisms of the functions of pseudoknots such as hTR.

Quantitative predictions for RNA secondary structures and secondary structure folding thermodynamics have been quite successful.[19–25] One of the possible limiting factors for predicting RNA folding has been the treatment of RNA pseudoknots. Recently, several computational models have been developed for pseudoknot structural prediction.[26–32] In particular, based on the rigorous statistical mechanics and a virtual bond model for pseudoknot structures, we recently calculated and tabulated pseudoknot loop entropy parameters for different pseudoknots. Using the loop entropy parameter table, we can predict free energy landscape for RNA pseudoknots, including the native structure and all the possible thermodynamic intermediates and their stabilities. However, despite the successful development of the folding thermodynamic stability and structural prediction models, there are relatively few folding kinetics models. Besides Monte Carlo[33] and analytical[34] kinetics calculations for RNA secondary structures, several computer simulation studies have been performed to analyze the dynamic behavior of hTR pseudoknot[35] and the folding pathway of hepatitis delta virus ribozyme.[36] These simulational studies provided useful insights into pseudoknot functions. For example, the simulated dynamic behaviors of hTR pseudoknot provided information about the structural fluctuations for hairpin with tandem mismatches and the related functional role.[35] However, the simulational studies are based on incomplete conformational sampling. The complete conformational ensemble for pseudoknot and secondary structures is huge. Monte Carlo simulations may miss some of the low-stability conformations, leading to incomplete conformational sampling. Such low-stability conformations can be important for the formation of transition states and possible kinetic traps. In the present study, we consider the complete conformational ensemble, including all the possible native-like and misfolded states. The work is motivated by the need to explore the global and complete landscape of the pseudoknot folding kinetics, such as the folding pathways for the experimentally proposed molecular switch of hTR.

Our pseudoknot folding kinetics model relies on the recently developed RNA pseudoknot folding thermodynamic model (denoted as *Vfold*, where V stands for Virtual Bond Model).[32] Based on the thermodynamic parameters derived from *Vfold*, we develop a combined master equation and kinetic cluster theory[34] for pseudoknot folding kinetics. From the model, we obtain folding pathways, inter-

mediates, rates, rate-limiting steps, activation energies, etc., for RNA pseudoknot. After developing the model, we validate the model through comparisons with extensive experimental data such as the Arrhenius plot and the folding rate for the PK5 pseudoknot (a synthetic oligonucleotide).[37,38] For a great variety of RNA pseudoknots, we find that pseudoknots fold/unfold through two-step (biphasic) processes, where a hairpin is formed as a kinetic intermediate in the first step followed by the rate-limiting disruption of the misfolded structures and the formation of the native pseudoknot. As a further application of the model, we investigate the kinetic pathways for the folding of the hTR pseudoknot.[10,11] The results are in accordance with the experiments. Moreover, the predicted kinetic pathways provide mechanisms for the proposed hairpin-pseudoknot molecular switch.

Model and Theory

Kinetic moves and master equation

The master equation is constructed from the transition rates between different states. We define a state by the base-pairing pattern. For a given sequence, we enumerate all the possible base-pair patterns (=states) for the secondary structures and the pseudoknotted structures. For each given state, the helical stems are modeled as an A-form RNA helix, while the unpaired segments of the chain, such as the loops, are treated as flexible chains that can adopt multiple conformations. Therefore, a state usually corresponds to many accessible conformations. For a given state, the energies (enthalpies) of the base-pairing/stacking are obtained from the empirical experimental parameters[39,40] and the entropy is calculated from the *Vfold* model.[23,24,32] In the *Vfold* model, the conformation of the helix stems is described by the atomic coordinates of the virtual bonds of A-form RNA helices, and the loop conformations are exhaustively enumerated through self-avoiding walks of the virtual bonds. The number of conformations for a given state is determined by the viable conformations of the loops and other non-base-paired regions, the different orientations of the helix stems, and the volume exclusion between different parts of the chain. An important feature of the *Vfold* model is that it takes into account the mismatched intraloop base-pairing/stacking interactions.

Since RNA helices are mainly stabilized by the base stacking force, we assume that different RNA conformations are inter-converted *via* the formation and disruption of a base stack such as (*a*, *b*, *c*, *d*) in Figure 1(c)-1 or a stacked base-pair such as (*x*, *y*) in Figure 1(c)-2.

To calculate the rate constant for a kinetic move, we need a transition state model. For the formation of a base-pair/stack, we assume that the transition state occurs when the participating nucleotides have been restricted to the full base-paired positions, but not yet stabilized by the hydrogen bonding/base stacking interactions. Therefore, the entropy of the transition state $S^{\ddagger}$ is equal to the entropy $S^{(\mathrm{close})}$ of the stable base-paired state and the enthalpy of the transition state $H^{\ddagger}$ is equal to the enthalpy $H^{(\mathrm{open})}$ of the unpaired open state. Since the base-pairing/stacking is short-ranged, we expect that the specific configuration of the transition state is in the close vicinity of the stable base-pairing configuration. The precise configuration of the transition state requires detailed atomic computation. According to the transition state assumption, the kinetic barrier for the formation of a base-pair is entropic, as characterized by the entropic loss ($S^{\ddagger}-S^{(\mathrm{open})}$) upon the formation of the base-pair/stack.

Similarly, for the disruption of a base-pair/stack, we assume that the transition state occurs when the stabilizing interactions have been disrupted (so $H^{\ddagger}\simeq H^{(\mathrm{open})}$) but the nucleotides are not yet liberated from the base-paired positions (so $S^{\ddagger}\simeq S^{(\mathrm{close})}$). Such a transition state would have the same configuration as the transition state for the formation of a base-pair/stack as defined above. The kinetic barrier for the disruption of a base-pair/stack is enthalpic, as quantified by the activation enthalpy ($H^{\ddagger}-H^{(\mathrm{close})}$) required to activate the transition from the stable base-paired state to the transition state. We note that the above transition state model has been used to compute RNA hairpin folding kinetics and has given good theory-experiment agreements.[41]

For a kinetic move between states N and C in Figure 1(c), the activation free energies are $\Delta G^{\ddagger}_{+}=T(S_C-S_N)$ and $\Delta G^{\ddagger}_{-}=H_C-H_N$ for the formation and disruption of a base-pair/stack, respectively. Here S_x and H_x denote the entropy and enthalpy of state x (=N or C), respectively. The corresponding rate constants are thus given by the following equations:

$$k_{+}=k_0\ e^{(S_N-S_C)/k_{\mathrm{B}}},\ k_{-}=k_0\ e^{(H_N-H_C)/k_{\mathrm{B}}T}$$

where k_{B} is the Boltzmann constant and T is the temperature. The constant k_0 is a prefactor fitted from the experimental data.[41] The rate constants for all transitions that cannot be connected by kinetic moves are set to be zero.

The time evolution of the population $p_i(t)$ for the ith state is governed by the following master equation:

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}=\sum_{j\neq i}^{\omega}\left[k_{j\to i}p_j(t)-k_{i\to j}p_i(t)\right] \qquad (1)$$

where ω is the total number of conformations and $k_{x\to y}$ is the transition rate from state x to state y. The rate constants constitute an $\omega\times\omega$ rate matrix $\mathbf{M}$, defined by matrix element: $M_{ij}=k_{j\to i}$ for $i\neq j$ and $M_{ii}=-\sum_{j\neq i}^{\omega}k_{i\to j}$. From the eigenvalues λ_m (m=1, 2,…, ω) and eigenvectors $\mathbf{n_m}$, we obtain the populational kinetics for any given initial conformational distribution:

$$p(\mathrm{t})=\sum_{m=1}^{\omega}C_m\mathbf{n_m}e^{\lambda_m t} \qquad (2)$$

where $p(\mathbf{t})=[p_1(t), p_2(t),\ldots, p_\omega(t)]$ is the populational vector and C_m's are the coefficients dependent on the initial condition.

The advantages of the master equation are that (a) its solution is analytical and (b) the solution is based on the complete conformational ensemble. The disadvantages are (a) it cannot treat large molecules (>30 nt) because of the large conformational space and (b) it cannot give microscopic pathways. Therefore, in order to treat long chains and to obtain the detailed pathways, we use the kinetic cluster approach, as described below.

Kinetic cluster method

For a system with discrete rate-limiting folding/unfolding steps, we can classify the conformational ensemble into preequilibrated clusters. Different clusters are separated by the rate-limiting (slow) steps.[42,43] The overall kinetics is determined by the much reduced network of clusters, from which we can perform the detailed kinetic analysis.[42]

From the inter-conformation transition rate, for a folding process, a rate-limiting step is either (a) the formation of native base-pairs (stacks) with large entropic decreases or (b) the disruption of the non-native base-pairs (stacks) with large enthalpic increases. For an unfolding process, rate-limiting steps are the disruptions of the native base-pairs (stacks) with large enthalpic increases.

For each pair of clusters C and C′ (see Figure 2), there usually exist multiple micro-pathways between them. Each micro-pathway connects a conformation i in cluster C to a conformation i′ in cluster C′. Throughout the study, we use micro-pathways to denote structural transformations at the level of conformations and use macro-pathways to denote transformations at the level of clusters. Each inter-cluster micro-pathway i→i′ between clusters C and C′ involves the rate-limiting kinetic move that separates the two clusters. The overall inter-cluster transition rates $k_{C\to C'}$ and $k_{C'\to C}$ are the averages over all the micro-pathways:

$$k_{C\to C'}=\sum_{\text{pathways}} p_i k_{i\to i'},\ k_{C'\to C}=\sum_{\text{pathways}} p_{i'}k_{i'\to i} \quad (3)$$

where p_x (x=i or i′) is the equilibrium fractional population of conformation x in the respective cluster.

Among all the micro-pathways from C to C′, the probability of folding or unfolding along the i→i′ micro-pathway is determined by the following partitioning probability (see equation (4)):

$$f_i^{\text{micro-path}}=p_i k_{i\to i'}/k_{C\to C'} \quad (4)$$

The largest $f^{\text{micro-path}}$ gives the dominant micro-pathway.

A cluster is usually connected to several other clusters. The probability for a molecule to escape from cluster C through the C→C′ transition is determined by the following fraction.

$$f_{C\to C'}^{\text{macro-path}}=k_{C\to C'}\Big/\sum k_{C\to\text{all other clusters connected to C}} \quad (5)$$

The largest $f^{\text{macro-path}}$ gives the dominant macro-pathway.

Folding rates with multiple intermediates

From the kinetic cluster method, we can identify the dominant macro-pathway. How to compute the rate constant along a macro-pathway, which, in general, can have multiple inter-cluster transitions (see Figure 2)? If there exists

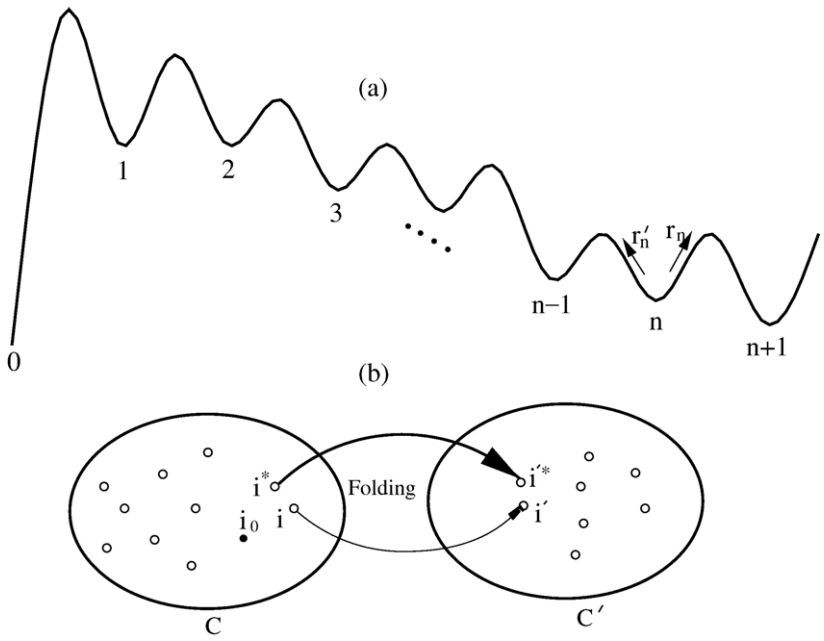

Figure 2. (a) A schematic pathway from the initial cluster (state) 0 to the final cluster (state) $n+1$ with n intermediate clusters (states). (b) Inter-cluster transition between clusters C and C′, where i→i′ denotes a micro-pathway for the inter-cluster transition, i*→i′* denotes the dominant micro-pathway, and i_0 denotes the most stable state in cluster C.

only one intermediate ($n=1$ in Figure 2(a)), the lowest eigenvalue of the rate matrix for the three-state (0, 1, 2 in Figure 2(a)) system gives the rate k_2 for $0\rightarrow1\rightarrow2$:

$$k_2 = -0.5\left(k_{0\rightarrow1}+k_{1\rightarrow0}+k_{1\rightarrow2}+k_{2\rightarrow1}+\sqrt{(k_{0\rightarrow1}-k_{2\rightarrow1})^2+(k_{1\rightarrow0}+k_{1\rightarrow2})^2+2(k_{1\rightarrow0}-k_{1\rightarrow2})(k_{0\rightarrow1}-k_{2\rightarrow1})}\right) \quad (6)$$

If the intermediate states (clusters) for a given macro-pathway are short-lived, we can then obtain an analytical expression for the rate constant as shown below. The condition of the short lifetime for the intermediates leads to the following recursive relation between rate k_n for pathway $0\rightarrow n$ and k_{n-1} for pathway $0\rightarrow n-1$ in Figure 2(a):

$$k_n = k_{n-1}r_{n-1}\sum_{p=0}^{\infty}(r_n' r_{n-1})^p = k_{n-1}\left(\frac{r_{n-1}}{1-r_n' r_{n-1}}\right) \quad (7)$$

where the sum corresponds to the iterative rebounce processes and r_{n-1} and r_n' are the forward and backward (rebounce) probabilities for a molecule in C_{n-1} and C_n, respectively:

$$r_n' = \frac{k_{n\rightarrow n-1}}{k_{n\rightarrow n+1}+k_{n\rightarrow n-1}}; \quad r_{n-1} = \frac{k_{n-1\rightarrow n}}{k_{n-1\rightarrow n}+k_{n-1\rightarrow n-2}}$$

The above recursive relation results in the rate for a macro-pathway with n intermediates:

$$k_n = k_1 r_{n-1}\prod_{p=1}^{n-2}\frac{r_p}{1-r_{p+1}' r_p} \quad (8)$$

RNA pseudoknot folding and unfolding often involve several parallel folding pathways. The total folding rate is the sum of the rate over all the parallel pathways.

Activation energy

The temperature dependence of the rate constant (Arrhenius plot) reveals the information about the microscopic mechanism of the folding/unfolding kinetics.[34,44–50] For simple two-state kinetics, the activation energy E_a (the change in the mean energy (enthalpy) in the rate-limiting step) can be extracted from the Arrhenius equation:

$$E_a = -d\ln k/d(1/k_B T) \quad (9)$$

If the folding process is rate-limited by an inter-cluster transition $C\rightarrow C'$, then equations (3) and (9) lead to the following expression for E_a:

$$E_a = \langle H\rangle_{\text{micro–path}} - \langle H\rangle_{\text{cluster}} + <\varepsilon_a>_{\text{micro–path}} \quad (10)$$

Here H is the enthalpy of a conformation, ε denotes the activation barrier $\varepsilon^{(i\rightarrow i')} = -d\ln k_{i\rightarrow i'}/d(1/k_B T)$ for micro-pathway $i\rightarrow i'$, and $<>_{\text{micro-path}}$ and $<>_{\text{cluster}}$ are the averages over the micro-pathways weighted by the fractional pathway partitioning factor $f_{i\rightarrow i'}^{\text{micro-path}}$ (equation (4)) and the conformations in cluster C weighted by the equilibrium Boltzmann factor, respectively:

$$\langle H\rangle_{\text{cluster}} = \sum_{i\epsilon C} p_i H_i,$$

$$\langle H\rangle_{\text{micro–path}} = \sum_{i\rightarrow i'} f_{i\rightarrow i'}^{\text{micro–path}} H_i,$$

$$<\varepsilon_a>_{\text{micro–path}} = \sum_{i\rightarrow i'} f_{i\rightarrow i'}^{\text{micro–path}} \varepsilon_a^{(i\rightarrow i')}$$

Clusters are separated by a rate-limiting step. $\varepsilon_a^{(i\rightarrow i')}$ is the activation barrier for such rate-limiting step. Therefore, $\varepsilon_a^{(i\rightarrow i')}$ is independent of the micro-pathway $i\rightarrow i'$ and is a constant ε_a. As a result, the activation barrier E_a has a much simpler form:

$$E_a = \langle H\rangle_{\text{micro–path}} - \langle H\rangle_{\text{cluster}} + \varepsilon_a \quad (11)$$

If $C\rightarrow C'$ is rate-limited by the formation/breaking of a base-pair (stack) whose kinetic barrier is entropic/enthalpic, ε_a would be zero/positive (=the enthalpic change for the disruption of the rate-limiting base stack).

To understand the physical meaning of the expression for E_a as shown in Figure 2(b), we use i_0 to denote the most stable conformation in cluster C and assume $i^*\rightarrow i'^*$ to be the most probable micro-pathway for $C\rightarrow C'$. Then the activation energy in equation (10) gives the enthalpic difference between i^* and i_0. Physically, this is the energy required to excite the transition from the most highly populated state i_0 to the kinetically most active state i^*.

As a test, we find that the number of conformations for RNA secondary structures and pseudoknotted structures grows with the chain length L as $\omega=(0.2)(1.35)^L$ for a 77-nt transfer RNA sequence.[51,52] Given the large number ω of the conformations, the most time-consuming step of the computational process is the calculation of the large number of the rate constants between all the possible kinetically connected pairs of the conformations. Therefore, the computation time for the folding kinetics calculation scales as $\omega\times\omega$ (=the number of conformational pairs). In the kinetic cluster approach, the number of clusters is dramatically smaller than the original number of conformations. Therefore, unlike the computation time, the limiting factor for the required computer memory in the calculation is not the inter-cluster rate matrix. Instead, the memory is determined by the required space to save the large number of the original conformations. Therefore, the required memory scales as ω. For a 45-nt sequence that we tested, there are $\omega=163{,}578$ secondary and pseudoknot conformations. The folding kinetic computation took 150 min with a 56 M memory.

Results and Discussion

PK5 pseudoknot

As shown in Figure 3, PK5 pseudoknot is a 26-nt synthetic oligonucleotide,[37,38] which consists of two stems connected by two loops. We choose PK5 because it is one of the few pseudoknots of which the folding kinetics, including the activation energy and unfolding rate, has been experimentally measured.

Using the *Vfold* model for RNA pseudoknot folding thermodynamics,[32] we predict the native structure and the thermal stability for PK5. One of the problems presented in the folding thermodynamic calculation is how to treat the energetics for mismatched base-pairs. For a single mismatched base-pair in the internal loop, we add entropy and enthalpy penalties according to the sequence-dependent energy parameters obtained from experiments.[38,40] Theory-experiment[38] comparison shows that our native structure prediction is exact, and the predicted stability −4.2 kcal/mol is very close to the experimental results of −4.3 kcal/mol (NMR) and −4.9 kcal/mol (optical).[38] Moreover, the predicted melting temperature, T_m=67 °C, is also close to the experimental result, 64 °C.[38]

Folding kinetics experiments show that the folding of PK5 involves a misfolded 5′ hairpin[38] as a kinetic intermediate. The 5′ hairpin contains three non-native base-pairs, which must be disrupted in order for the native stem 1 to be formed (see Figure 3). We call a base-pair to be "native" if it exists in the native structure (PK5 pseudoknot) and "non-native" otherwise. Our folding thermodynamic calculation shows that the 5′ hairpin is substantially less stable than the native pseudoknot PK5 with an equilibrium constant [PK5]/[5′ hairpin]=13.1 at T=37 °C. Therefore, the 5′ hairpin is thermodynamically unstable and can only possibly emerge in the form of a transient intermediate state.

Wyatt *et al.* have measured the (pseudoknot)→(5′ hairpin) unfolding rates at five different temperatures: 19.4 °C, 22.2 °C, 24.5 °C, 27.2 °C, and 29.5 °C. The Arrhenius plot for the temperature dependence of the unfolding rate gives an apparent activation energy of about 42 kcal/mol. The experiment study suggested that the unfolding activation barrier corresponds to the disruption of stem 2 in the PK5 pseudoknot.[38] Based on the kinetic cluster analysis described above, we can predict not only the experimentally measured activation energy and the rate constants but also the detailed folding pathways for PK5 pseudoknot (Table 1).

Relaxation kinetics

By exhaustively enumerating all the different arrangements of the base stacks, we find that the 26-nt pseudoknot-forming sequence can form 1328 possible secondary structures and 177 pseudoknotted conformations, with a total of 1505 states. We label the conformation state using $I_{(N,NN)}$, where N is the number of the native base-pairs and NN is the number of non-native base-pairs. The two stems of a pseudoknot may or may not form coaxial stacking. For the PK5 pseudoknot, experimental results suggest the possibility of partial coaxial stacking.[37] Our predicted folding stability without coaxial stacking shows a better experimental agreement than that with coaxial stacking,[32] indicating that the coaxial stacking is weak. Therefore, as a

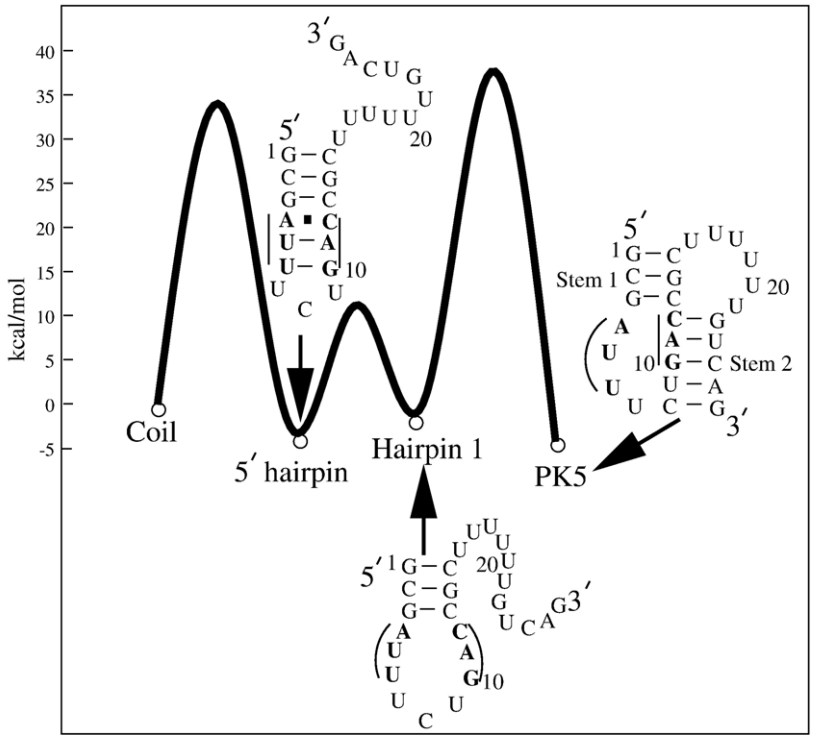

Figure 3. A schematic free energy landscape for the transition from the unfolded (coil) state to the PK5 pseudoknot at T=37 °C. A three-state process has been observed by the experiment.[38] The single-strand coil is found to form the misfolded hairpin first followed by the disruption of the three misfolded base-pairs (in bold). Then hairpin 1 folds into the native pseudoknot (PK5).

Table 1. The thermodynamic parameters for the loop formation with single mismatches

Sequences	ΔH_{exp}	ΔS_{exp}	ΔH_{cal}	ΔS_{cal}	$\Delta\Delta H$	$\Delta\Delta S$
5′GA(GCG)AG3′/3′CU(CAC)UC5′	−13.4	−41.6	−16.7	−43.9	3.3	2.3
5′GA(GAG)AG3′/3′CU(CGC)UC5′	−11.4	−36.5	−15.0	−35.2	3.6	−1.3
5′GA(GCG)AG′3/3′CU(CUC)UC5′	−10.4	−33.6	−11.3	−30.0	0.9	−3.6
5′GA(GUG)AG3′/3′CU(CUC)UC5′	−12.6	−40.6	−14.3	−40.0	1.7	−0.6
5′GA(GAG)AG3′/3′CU(CAC)UC5′	−12.3	−40.9	−15.5	−41.9	3.2	1.0
5′GA(GUG)AG3′/3′CU(CCC)UC5′	−7.7	−26.4	−11.7	−32.6	4.0	6.2
5′GC(GAU)U3′/3′CG(CCA)G5′	−10.2	−29.4	−11.4	−30.3	1.2	0.9
Average	–	–	–	–	2.6	0.7

Sequences in lines 1 to 6 are from Kierzek *et al.*[40] The last sequence is from Wyatt *et al.*[38]

simplified approximation, we assume no coaxial stacking for PK5. Solving the eigenvalues and eigenvectors of the 1505×1505 matrix requires 6 min on a 3.06-GHz PC.

The 1505×1505 rate matrix for the complete conformational ensemble gives the full populational kinetics for each and every state. In Figure 4, we show the populational kinetics in the folding process from the completely unfolded state at temperature T=19.4 °C. The populational kinetics curves clearly show that, consistent with the experiment,[38] the misfolded 5′ hairpin (denoted as $I_{(3,3)}$) is a kinetic intermediate.

In order to obtain the detailed information about folding pathways, rates, and rate-limiting steps, we go beyond the conformational ensemble-based master equation approach by using the kinetic cluster approach. From the exhaustive examination for all the possible conformations, we find five rate-limiting base stacks. Four native base stacks, 10GA11-23UC24, 1GC2-14GC15, 9UG10-24CA25, and 11AC12-22GU23, are slow to form because of the large entropic decreases of ΔS=−35.5, −34.9, −27.9, and −26.2 kcal/mol·K, respectively. The formation of these four base stacks is thus rate-limiting. In addition, the disruption of the non-native base stack 3GA4-23UC24 is also slow due to the large enthalpic increase of ΔH=13.3 kcal/mol in the disruption process and is hence a rate-limiting base stack. According to the formation/disruption of these five base stacks, we classify the original conformational ensemble into 15 kinetic clusters. Conformations that do not contain any of these rate-limiting base stacks are classified as cluster L_U, the misfolded 5′ hairpin (see Figure 3) is in a cluster denoted as L_1, and the native PK5 pseudoknot is in the native cluster L_N.

The overall folding kinetics can be described by the 15×15 rate matrix. The lowest nonzero eigenvalue of the rate matrix, denoted as λ_1, corresponds to the rate constant of the slowest kinetic mode. If there exists a single pronounced slowest mode (smallest λ_1), then the kinetic process is single exponential and the relaxation rate k_r is equal to λ_1. For the system under folding condition ($T<T_m$=67 °C), the relaxation rate k_r is approximately equal to the folding rate k_f. Likewise, under the unfolding condition ($T>T_m$), it is approximately equal to the unfolding rate k_u.

In Figure 5(a), we show the temperature dependence of λ_1. We find that the eigenvector of the slowest mode is dominated by the components corresponding to the coil state and the native pseudoknot, suggesting that the slowest model for the folding process is the transition from the coil state to the native pseudoknot. The comparison with the

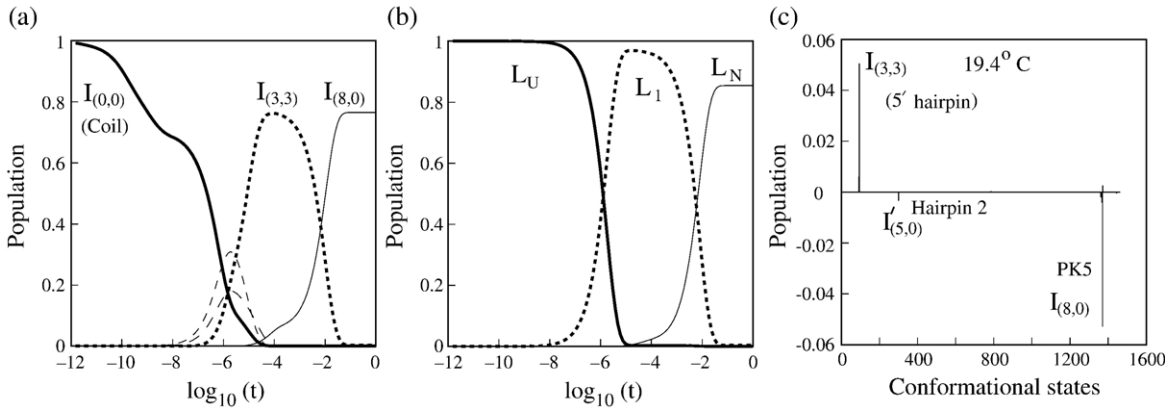

Figure 4. Folding kinetics for PK5 at T=19.4 °C. Populational kinetics from the exact master equation (a) and the kinetic cluster method (b). Populational curves for states whose populations never reach 10% are not shown. The two methods unambiguously show that the misfolded state ($I_{(3,3)}$), which is in cluster L_1, is a kinetic intermediate in the folding from the coil state ($I_{(0,0)}$) to the native pseudoknot ($I_{(8,0)}$). The two states whose populational kinetics are shown as thin dashed lines in (a) are contained in cluster L_U. (c) The eigenvectors for the slowest mode (λ_1). Hairpin 2 is a stem–loop structure with stem 2 closed by a hairpin loop.

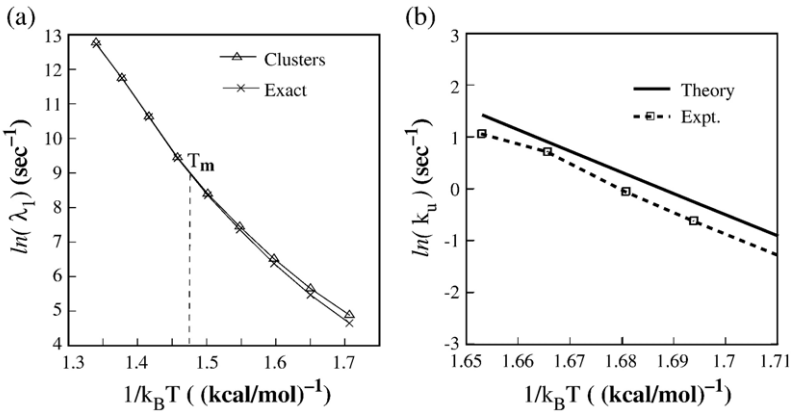

Figure 5. (a) The Arrhenius plot for the lowest nonzero eigenvalue from the exact master equation and the kinetic cluster method. (b) The comparison of the unfolding rates between the experiments (dot line) and the calculations. The unfolding rates are measured at five different temperatures: 19.4 °C, 22.2 °C, 24.5 °C, 27.2 °C, and 29.5 °C. The unfolding process is from the native pseudoknot ($I_{(8,0)}$) to a misfolded 5′ hairpin ($I_{(3,3)}$).

exact master equation results (Figure 5(a)) shows that the 15 clusters can well describe the folding kinetics.

Folding pathways (19 °C–30 °C)

In the experiment, the rate constants are measured in the temperature range between 19 °C and 30 °C.[38] Our kinetic cluster theory gives two apparent parallel folding pathways. As shown in Figure 6, the two folding pathways are $L_U \rightarrow L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_N$ (pathway 1) and $L_U \rightarrow L_1' \rightarrow L_2' \rightarrow L_3' \rightarrow L_N$ (pathway 2). Also shown in the figure are the dominant micro-pathways between the kinetically adjacent clusters on each pathway. We find that the folding pathways can be well described as biphasic processes, as explained in the following.

Phase I: Fast trapping in the misfolded state (10^{-5}–10^{-3} s). The coil state ($I_{(0,0)}$) in cluster L_U rapidly folds into a misfolded 5′ hairpin ($I_{(3,3)}$ in cluster L_1). Because the detrapping rate from the misfolded state is slow, the misfolded 5′ hairpin population would quickly accumulate to form a kinetic trap. Figure 4(a) shows that about 80% of the population is trapped in the misfolded state $I_{(3,3)}$. What causes the fast folding from the coil state to the misfolded 5′ hairpin $I_{(3,3)}$? First, from Figure 6, we find that hairpin $I_{(2,3)}$ is the most populated state in the preequilibrated distribution of cluster L_U (25.2%). With a single kinetic move (the formation of the (1, 2, 14, 15) base stack), $I_{(2,3)}$ is rapidly converted into the misfolded state $I_{(3,3)}$. The pathway from cluster L_U to L_1' is less probable because the pathway conformation $I'_{(2,3)}$ in cluster L_U is much less populated (0.03%) than $I_{(2,3)}$ (25.2%). Equation (3) gives that $k_{L_U \rightarrow L_1} = 5.4 \times 10^5\ s^{-1} \gg k_{L_U \rightarrow L_1'} = 1.4 \times 10^4\ s^{-1}$. As a result, most population would be trapped in the misfolded state $I_{(3,3)}$ at $t \sim (k_{L_U \rightarrow L_1})^{-1} \sim 10^{-5}$ s, which agrees with the populational kinetics curve shown in Figure 4(a) and (b).

Phase II: Slow detrapping from the misfolded state (10^{-3}–10^{-1} s). From Figure 6, we find that the slowest step for the overall kinetics is the detrapping process $L_1 \rightarrow L_2$. Since the most stable state in cluster L_1 is the misfolded 5′ hairpin $I_{(3,3)}$, the detrapping transition from L_1 corresponds to the breaking of $I_{(3,3)}$. There exist two detrapping pathways from L_1: $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_N$ (pathway 1) and $L_1 \rightarrow L_U \rightarrow L_1' \rightarrow L_2' \rightarrow L_3' \rightarrow L_N$ (pathway 2), with rates (from equation (8)) $k_f^{(path1)} = 64.0\ s^{-1}$ and $k_f^{(path2)} = 33.1\ s^{-1}$, respectively. Therefore, the total folding rate (k_F) can be calculated as a sum of the two pathways $k_F = k_{L_1 \rightarrow L_N} \simeq k_f^{(path1)} + k_f^{(path2)} = 97.1\ s^{-1}$, which is close to $100.1\ s^{-1}$ from the exact master equation. The estimated rate, which accounts for the contributions only from the above two dominant pathways, is slightly smaller than the total folding rate from the cluster analysis. This is due to the neglected contributions from other pathways. Because $k_f^{(path1)} \gg k_f^{(path2)}$, the majority of the population in cluster L_1 would fold along pathway 1 into the final native state. Therefore, pathway 1 is the most dominant folding pathway. In the overall folding process, the transition from cluster L_1 to cluster L_2, i.e., the disruption of the misfolded hairpin $I_{(3,3)}$, is rate-limiting.

Activation barrier of folding (19 °C–30 °C)

At temperature 19.4 °C, the main folding pathway is $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_N$ in Figure 6. The slowest (rate-limiting) inter-cluster transition is $L_1 \rightarrow L_2$, which has a rate constant of $72.2\ s^{-1}$. Equation (10) gives the activation energy for $L_1 \rightarrow L_2$ as $E_a^{(path1)} = \langle H \rangle_{micro\text{-}path} - \langle H \rangle_{cluster\ L_1} = (-20.8) - (-33.9) = 13.1$ kcal/mol.

To understand the result, we note that the misfolded 5′ hairpin (state $I_{(3,3)}$) is the most stable state in cluster L_1 with a fractional population of 84%, while the dominated micro-pathway for $L_1 \rightarrow L_2$ is $I_{(3,0)} \rightarrow I_{(5,0)}$ with a fractional pathway partitioning factor of $f^{micro\text{-}path} \simeq 93\%$. Therefore, the activation energy can be estimated as the energy required to excite the molecule from $I_{(3,3)}$ to $I_{(3,0)}$. The two states differ by three base-pairs with a total enthalpy difference of 13.8 kcal/mol, which is close to $E^{(path1)} = 13.1$ kcal/mol.

For the other detrapping pathway ($L_U \rightarrow L_1' \rightarrow L_2' \rightarrow L_3' \rightarrow L_N$) from cluster L_1, which contains the 5′ hairpin $I_{(3,3)}$ as the most stable state, the rate-limiting step and the activation energy are determined by the $L_1 \rightarrow L_U$ transition, which is the slowest process on the pathway. The $L_1 \rightarrow L_U$

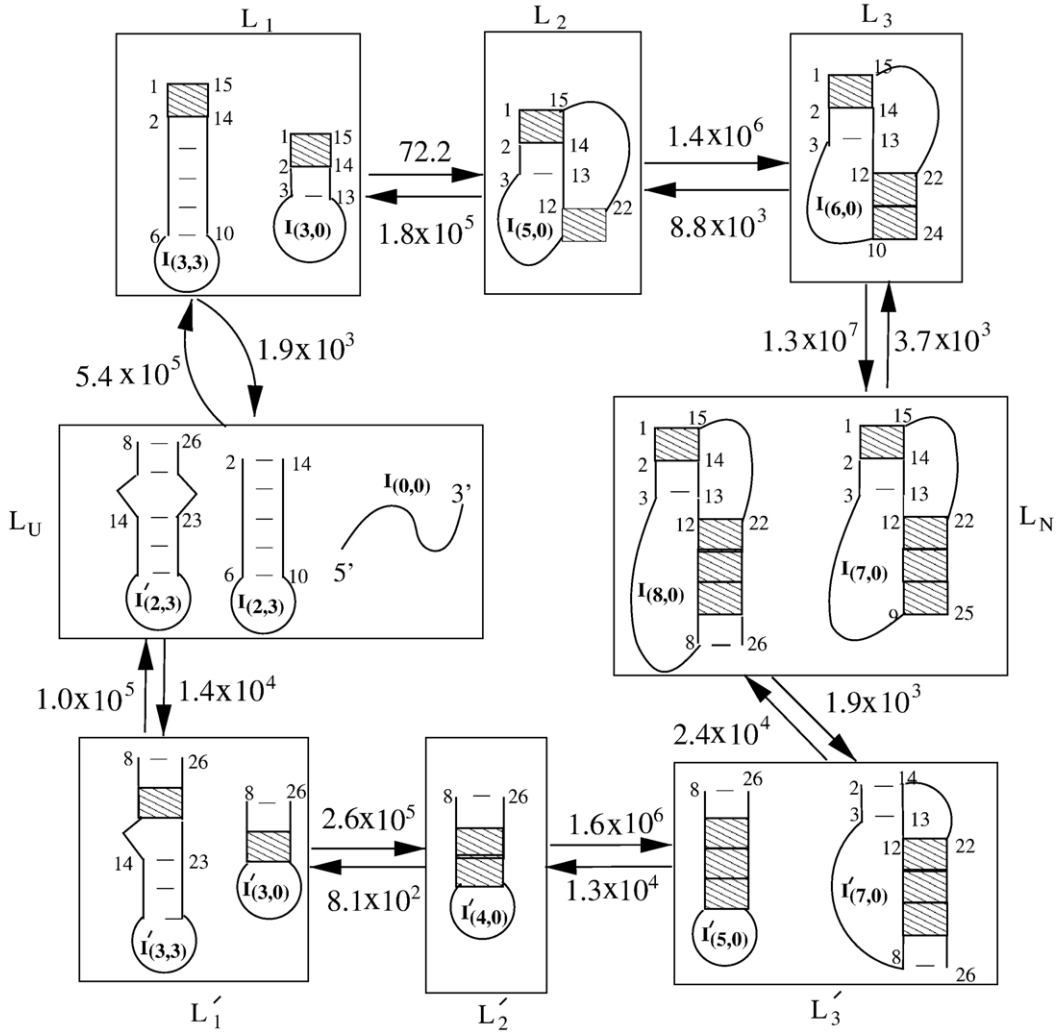

Figure 6. The folding of PK5 at temperature 19 °C–30 °C involves two parallel folding pathways: $L_U \rightarrow L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_N$ and $L_U \rightarrow L_1' \rightarrow L_2' \rightarrow L_3' \rightarrow L_N$. Also shown in the figure are the inter-cluster transition rates (in s^{-1}) and the dominant micro-pathways between adjacent clusters on the pathways. The shaded regions denote the rate-limiting base stacks.

transition is an unzipping process. Equation (10) gives the activation energy of $E_a^{(path2)} = 12.9$ kcal/mol, which is close to the enthalpic change of 14.2 kcal/mol for the breaking of the rate-limiting base stacking 1GC2-14GC15.

The overall activation energy of folding can be estimated as the following:

$$\begin{aligned} E_a &= \frac{k_f^{(path1)}}{k_f^{(path1)} + k_f^{(path2)}} E_a^{(path1)} \\ &\quad + \frac{k_f^{(path2)}}{k_f^{(path1)} + k_f^{(path2)}} E_a^{(path2)} \\ &= (66\%)(13.1) + (34\%)(12.9) \\ &= 13.0(\text{kcal/mol}) \end{aligned} \tag{12}$$

where $k_f^{(path2)}$ and $k_f^{(path2)}$ are the folding rates of pathway 1 and pathway 2, respectively. The result agrees exactly with that obtained from the slope of the Arrhenius plot in Figure 5(a).

Unfolding kinetics (19 °C – 30 °C)

From the eigenvector of the slowest mode (=relative populational changes of different conformations) in Figure 4(c), we find that the kinetic process can be primarily described as a transition $I_{(3,3)} \rightarrow I_{(8,0)}$, where the population of the misfolded conformation $I_{(3,3)}$ is depleted and most of the population is converted to the native pseudoknot $I_{(8,0)}$.

Because cluster L_3', which contains $I'_{(5,0)}$, and the native cluster L_N (see Figure 6) can inter-convert rapidly, L_3' and L_N can quickly preequilibrate to form a "macrostate." As a result, the overall kinetics is effective between a macrostate ($L_3' + N$) and the misfolded cluster L_1. For such a two-state transition, the detailed balance principle relates the unfolding rate k_U to the folding rate k_F:

$$k_U / k_F = \sum_{i \in L_1} e^{-\Delta G_i / k_B T} \Big/ \sum_{i \in (L_3' + N)} e^{-\Delta G_i / k_B T} \tag{13}$$

Here k_F under the folding condition (19 °C–30 °C) is approximately equal to the relaxation rate λ_1 (=the lowest eigenvalue of the rate matrix). In Figure 5(b), we show the theory-experiment comparison for the unfolding rate k_U at different temperatures. From the Arrhenius equation (equation (9)), the calculated unfolding rate gives an activation energy of 42 kcal/mol, which is in exact agreement with the experiment.[38]

The calculated unfolding rate is slightly faster than the experiment. In our model, we have neglected the additional stability of the pseudoknot due to partial coaxial stacking between helices[37] and the stabilizing tertiary interaction between loop and helices.[3,6,7] The inclusion of such effects would lower the unfolding rate. But these additional interactions may not cause a notable change in the activation energy (42 kcal/mol).

Unfolding kinetics ($T > T_m = 67$ °C)

At high temperature ($T = 80$ °C), the rate-limiting step for the unfolding of the native pseudoknot is controlled by the disruption of the base stacks of high enthalpic barriers. There exist two such rate-limiting base stacks, 1GC2-14GC15 and 10GA11-23UC24, whose disruption involves large enthalpic costs of $\Delta H = 14.2$ kcal/mol and 13.3 kcal/mol, respectively. According to the two rate-limiting base stacks, we classify the 1505 conformations into four clusters: cluster H_N for conformations with base stacks 1GC2-14GC15 and 10GA11-23UC24, cluster H_1 for conformations with base stack 1GC2-14GC15, cluster H_1' for conformations with base stack 10GA11-23UC24, cluster H_U for conformations without any of the two stacks.

As shown in Figure 7, the kinetic cluster method gives two parallel unfolding pathways: $H_N \rightarrow H_1' \rightarrow H_U$ (pathway 1) and $H_N \rightarrow H_1 \rightarrow H_U$ (pathway 2). Equation (6) gives the rate constants for each pathway: $1.7 \times 10^4\ s^{-1}$ for pathway 1 and $5.8 \times 10^4\ s^{-1}$ for pathway 2, respectively. Pathway 2 is likely a more probable unfolding pathway. On pathway 2, the pseudoknot unfolds from stem 1 followed by the breaking of stem 2. Furthermore, from Figure 7, we find that the rate-limiting step for the unfolding is from clusters H_1' to H_U, which, according to equation (10), has the activation energy of 27.1 kcal/mol.

The estimated activation energy is close to the slope of the Arrhenius plot in Figure 5(a) (29 kcal/mol). The positive activation is mainly from the disruption of three base stacks between states $I'_{(5,1)}$ and $I_{(2,1)}$. The total enthalpy for the three base stacks is 31.4 kcal/mol.

Human Telomerase RNA (hTR) pseudoknots

It has been proposed that the conformational switch between a hairpin structure, denoted as HP_{WT}, and the pseudoknot, denoted as PK_{WT}, is critical to hTR activity. One of the experimental evidence is that mutations destabilizing the hairpin or the pseudoknot have been found to cause hTR to lose activity.[9–11] On the other hand, a recent functional analysis of hTR[13] showed that the destabilization of the hairpin HP_{WT} could preserve the telomerase activity, which appeared to disagree with the conclusion about the functional importance of the hairpin ↔ pseudoknot conformational switch. We use an experimentally studied truncated hTR sequence[10] with 48 nt to study the folding kinetics of

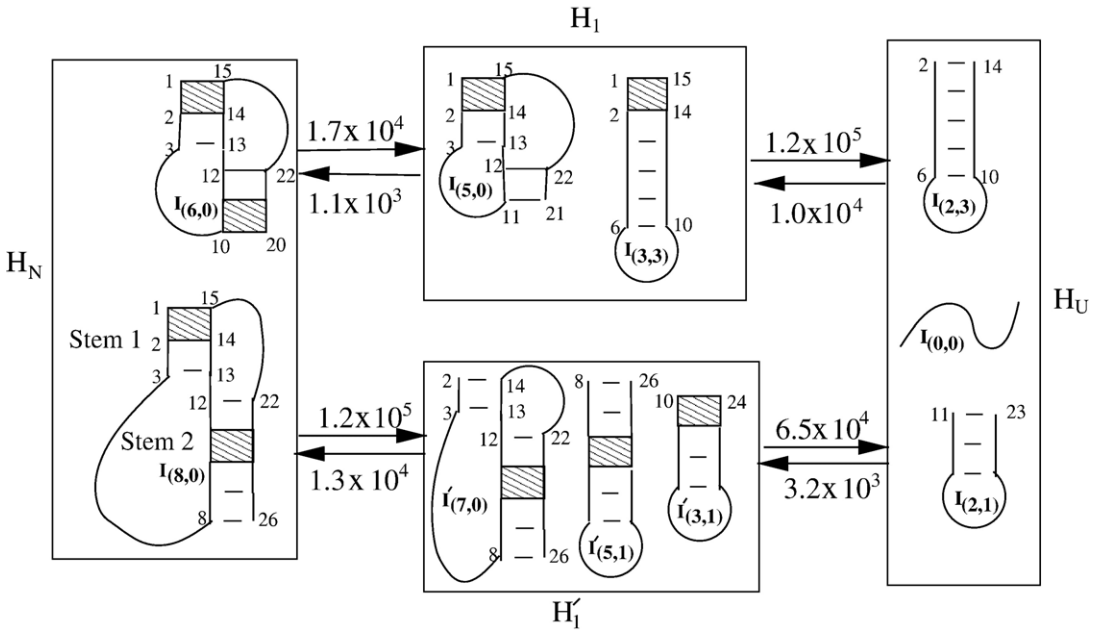

Figure 7. The unfolding pathway from the native PK5 pseudoknot to the coil state $I_{(0,0)}$. We find two parallel unfolding pathways ($H_N \rightarrow H_1 \rightarrow H_U$ and $H_N \rightarrow H_1' \rightarrow H_U$). Also shown in the figure are the inter-cluster rate constants (in s^{-1}). The rate-limiting base stacks are shown as the shaded regions.

hTR pseudoknot. The sequence contains two helix stems (p2b and p3) and the j2b/3 loop as well as the highly conserved nucleotide sequence of the j2a/3 loop in the hTR pseudoknot.[53]

We show the HP_{WT} hairpin structure in Figure 8(a). Thermodynamic experiments[10,11] suggest that the tandem mismatches in HP_{WT} can stabilize the structure. To extract the energy parameter for the tandem mismatches from the measured free energy ($\Delta G = -9.8$ kcal/mol) and enthalpy ($\Delta H = -92.0$ kcal/mol) for the hairpin,[10,11] we assume that three base stacks from base-pair (99U:U115) to (102U:C112) have approximately the same (ΔH, ΔS). With such approximation, the experimental results for (ΔG, ΔH) for the hairpin give (ΔH, ΔS) = (-1.8 kcal/mol, -9.2 kcal/mol K) for each base stack in the tandem mismatch. The other mismatches, 98GU-UC116, 102UU-AC112, and 104CU-UG110, are assigned the energy parameters according to the Turner rule.[39] To test the assigned energy parameters for the tandem mismatches, we calculate the heat capacity for the HP_{WT} and compare it with the experiment in Figure 8(b). The optical denaturation curves in the experiment are performed at 260 nm and 280 nm wavelengths. Consistent with the experiment, the calculated melting curve reveals two transitions in the experiment. The first peak corresponds to the unfolding of six base stacks from base-pair (99U:U115) to (105U:U109). And the second peak corresponds to the unfolding of six base stacks from base-pair (93G:C121) to (99U:U115). In general, the melting temperature in the calculation in 1 M NaCl is higher than that of the experiment in 200 mM due to lower ion concentration in the experiment.

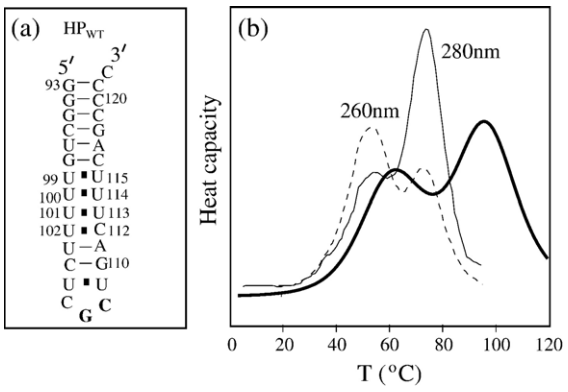

Figure 8. (a) The secondary structure of the hairpin-forming sequence HP_{WT}. (b) Comparison between the calculated heat capacity (solid line) in 1 M NaCl and the experimental optical spectroscopy at 260 nm and 280 nm in 200 mM KCl. Our calculation can well reproduce the two peaks in the experiments. The first peak corresponds to the disruption of the tandem mismatches from base-pair 99U:U115 to 105U:109U. The second peak corresponds to the disruption of the upper helix stem from base-pair 93G:C121 to 99U:U115. The calculated melting curve has been normalized according to the first peak in the experimental optical spectroscopy at 280 nm.

Folding kinetics

In Figure 9, we show the native structures of the wild-type pseudoknot-forming hTR sequence PK_{WT} and the experimentally studied mutants: $PK_{179AG/110CU}$,[13] PK_{DC}, and $PK_{\Delta U177}$.[9–11] In the figure, the native structures of PK_{WT}, $PK_{179AG/110CU}$, and $PK_{\Delta U177}$ are denoted by $I_{(15,0)}$ or N and the native structure of PK_{DC} is denoted by $I_{(6,7)}$ or C_1. For the $PK_{179AG/110CU}$ mutant, the double mutation 179AG/110CU destabilizes the hairpin HP_{WT} by disrupting the tandem mismatches in HP_{WT}. For the PK_{DC} mutant, the 107GC → AG mutation disrupts the most stable base stacking (5′GC-GC3′) in PK_{WT}, thus destabilizing PK_{WT}. In contrast, for $PK_{\Delta U177}$, the deletion of 177U eliminates the single bulge loop in PK_{WT}, thus stabilizing the pseudoknot PK_{WT}.

For PK_{WT}, $PK_{179AG/110CU}$, PK_{DC}, $PK_{\Delta U177}$, and the truncated short sequence HP_{WT}, we find 150,090, 77,335, 195,640, 133,569, and 3938 conformations, respectively. For HP_{WT}, we can use the exact master equation with a 3938×3938 rate matrix to obtain the exact solution for the eigenvalues and eigenvectors. However, for other sequences, the large number of conformations makes the computation with the master equation impossible, thus we use the kinetic cluster method.

In the kinetic cluster method, for PK_{WT}, we exhaustively examine the rate for the formation/disruption of each native and non-native base stack and classify the 150,090 conformations into 405 clusters. The eigenvalues and eigenvectors for the 405×405 rate matrix for the ensemble of clusters give the detailed folding kinetics for the PK_{WT} pseudoknot.

In Figure 10(a), we show the population kinetics of PK_{WT}. In the folding process, clusters C_1, C_1', and C_2 emerge as populated intermediates. In Figure 9, we show the kinetic pathways for PK_{WT}. We use the most stable structure in each cluster to represent the respective cluster. For example, the hairpin $I_{(6,7)}$ in the figure is the most stable structure in cluster C_2 for the PK_{WT} folding pathway. In the first step, this misfolded hairpin emerges as a kinetic intermediate. In the second step, the hairpin is unzipped to form the native pseudoknot $I_{(15,0)}$. Our kinetic folding pathway gives strong support for the formation of a transient hairpin structure and the conformational switch between the hairpin structure and the pseudoknot.[9,10] The hairpin structure $I_{(6,7)}$ ($=HP_{WT}$) and the native pseudoknot $I_{(15,0)}$ ($=PK_{WT}$) are formed in different stages of the folding pathway and can play different functional roles in telomerase activity.

Molecular switch and its relationship with telomerase activity

From the experiments,[9,10] it was found that the pseudoknot-destabilizing mutation PK_{DC} and the pseudoknot-stabilizing mutation $PK_{\Delta U177}$ can reduce the telomerase activity to a much lower level. It was proposed that a molecular switch

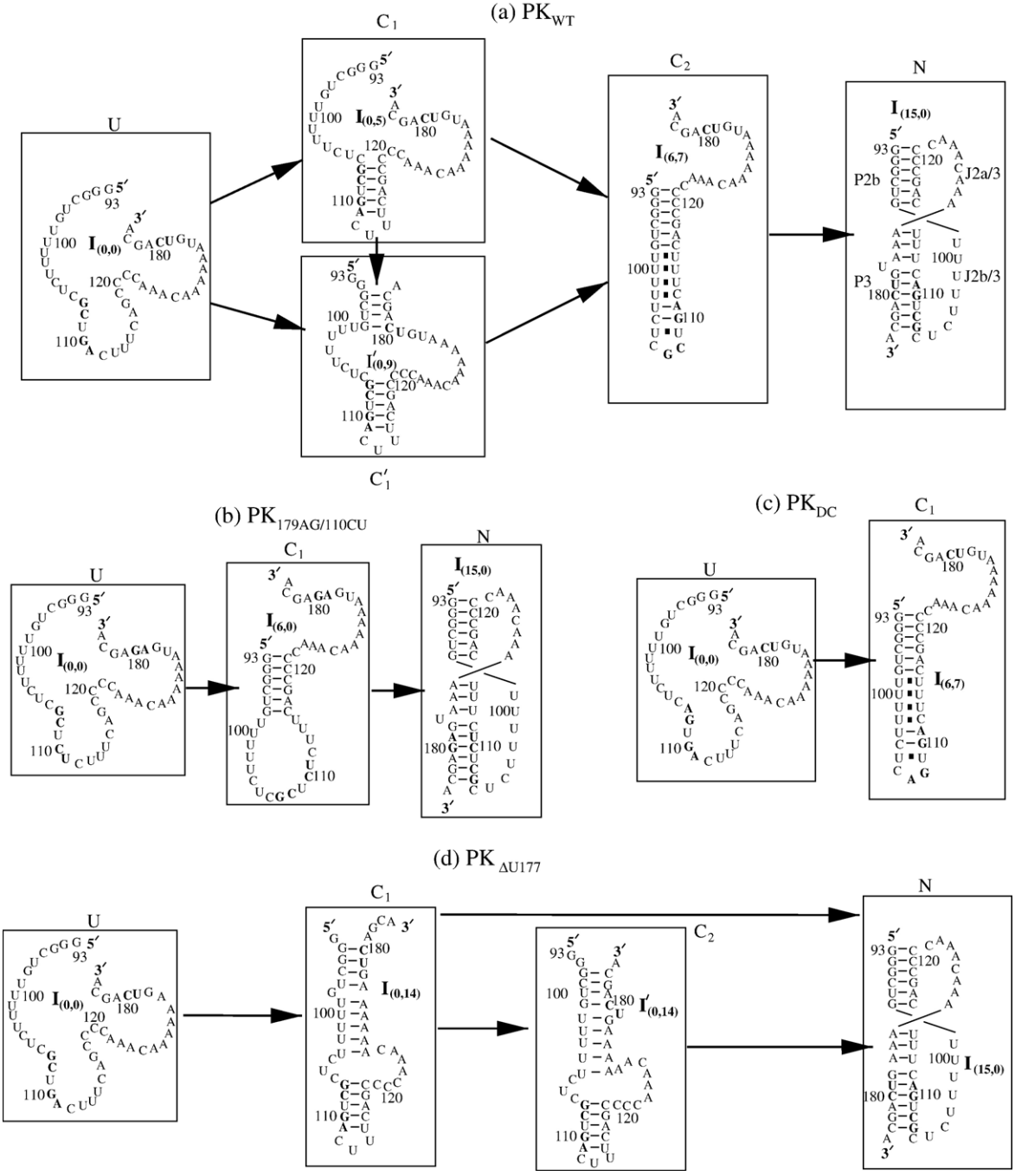

Figure 9. The kinetic intermediates appearing in the folding pathway for (a) PK_{WT}, (b) $PK_{179AG/110CU}$, (c) PK_{DC}, and (d) $PK_{\Delta U177}$ at $T = 37$ °C. We find that, for the mutants PK_{WT} and $PK_{179AG/110CU}$, the hairpin structure ($I_{(6,7)}$ in PK_{WT} and $I_{(6,0)}$ in $PK_{179AG/110CU}$) appears as a long-lived stable kinetic intermediate before the formation of the native pseudoknot. The native pseudoknot and the hairpin kinetic intermediate are absent for PK_{DC} and $PK_{\Delta U177}$, respectively.

between hairpin and pseudoknot may be important for the telomerase activity because the two mutations that alter the thermodynamic equilibrium between the hairpin and the pseudoknot result in the reduction in telomerase activity. However, in a different recent experiment,[13] a hairpin-destabilizing mutation ($PK_{179AG/110CU}$) was found to preserve the telomerase activity. How can the seemingly contradictory experimental results be consistent with each other?

We propose that the hTR activity is kinetically (instead of thermodynamically) controlled. The activity-preserving $PK_{179AG/110CU}$ mutation, though it destabilizes the hairpin state in thermal equilibrium, can actually cause the formation of the hairpin structure as a transient (kinetic) intermediate in the folding process. It is the switch between such kinetically formed (non-equilibrium) hairpin and the native pseudoknot that plays a role in telomerase activity.

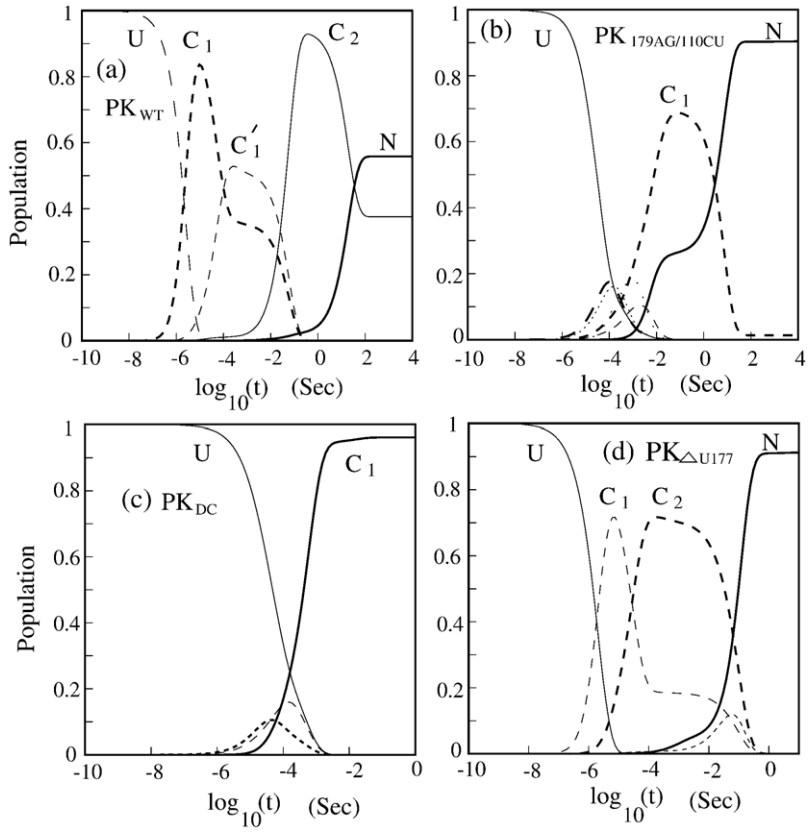

Figure 10. The populational kinetics for (a) the wild-type pseudoknot (PK_{WT}) and the three mutants (b) $PK_{179AG/110CU}$, (c) PK_{DC}, and (d) $PK_{\Delta U177}$ at $T = 37$ °C. We only show the cluster with more than 10% of the population's accumulation in the folding process.

To confirm the above hypothesis, we investigate the kinetic folding pathway for the three experimentally studied mutants. The original large conformational ensemble can be classified into 650, 434, and 290 clusters for $PK_{179AG/110CU}$, PK_{DC}, and $PK_{\Delta U177}$, respectively. In Figure 10(b), (c), and (d), we give the populational kinetics for the three mutants. The kinetic intermediates on the folding pathways are shown in Figure 9(b), (c), and (d). For $PK_{179AG/110CU}$, we find that, as we predicted, hairpin structure ($I_{(6,0)}$ in cluster C_1; see Figure 9(b)) is indeed a transient kinetic intermediate. Moreover, we find that the pseudoknot is absent in PK_{DC}, since the mutant destabilizes the pseudoknot stability by disrupting the two canonical base-pairs in stem 2. We note that experimental studies on PK_{DC}[10] suggested a partial formation (50% fractional population) of the pseudoknot. The theory-experiment disagreement is due to the neglected loop–stem tertiary interactions, which can significantly stabilize the PK_{DC} pseudoknot.[11] In addition, we find that the hairpin structure ($I_{(6,7)}$ or $I_{(6,0)}$) is absent in the folding pathway of $PK_{\Delta U177}$. The absence of the hairpin in $PK_{\Delta U177}$ and the pseudoknot in PK_{DC} on the kinetic pathways explains why both PK_{DC} and $PK_{\Delta U177}$ are not functionally active. From our kinetic study, we find that a stable pseudoknot in the equilibrium state (final state) and a kinetic hairpin intermediate may be essential for the telomerase activity. Such conclusion is consistent with the functional analysis of the telomerase activity.[9–11,13]

What causes the hairpin structure to be formed as a kinetic intermediate on the folding pathways for PK_{WT} and $PK_{179AG/110CU}$? The underlying mechanism is due to the slow folding rate for the pseudoknot. In Figure 11, we give the detailed folding pathway from the hairpin to the pseudoknot for $PK_{179AG/110CU}$. The folding rate for forming the pseudoknot is determined by two factors: (1) slow forward folding $C_1 \rightarrow C_2$ and (2) the competition between the rapid backward transition $C_3 \rightarrow C_2$ and the slow forward folding $C_3 \rightarrow C_4$. The $C_3 \rightarrow C_4$ transition corresponds to the formation of the single-nucleotide (U177) bulge loop and the AG:CU stack. Thus, deleting U177 as in mutant $PK_{\Delta U177}$ would accelerate the formation of the pseudoknot and thus prevent the formation of the hairpin C_1 in Figure 11 as a kinetic intermediate.

In conclusion, for the recent hairpin-destabilizing mutational experiment, we predict that the thermodynamically unstable hairpin structure ($I_{(6,0)}$) is formed as a kinetic intermediate.[13] This finding is consistent with the proposed hairpin-pseudoknot molecular switch mechanism for hTR activity.[9–11,13] Recently, Yingling and Shapiro performed molecular dynamics simulations to investigate the dynamics of the conformational fluctuations/changes for hTR hairpin[35] and pseudoknot.[54] In particular, the simulational studies predicted that the bulged U177 residue in the P3 helix stem of the pseudoknot (see the predicted structure *N* in Figure 9(a)) can form dynamic wobble U:U base-pairs with the uridine residues in the J2b/3 loop. Such U:U pairs would juxtapose the J2b/3 loop and the P3 stem,

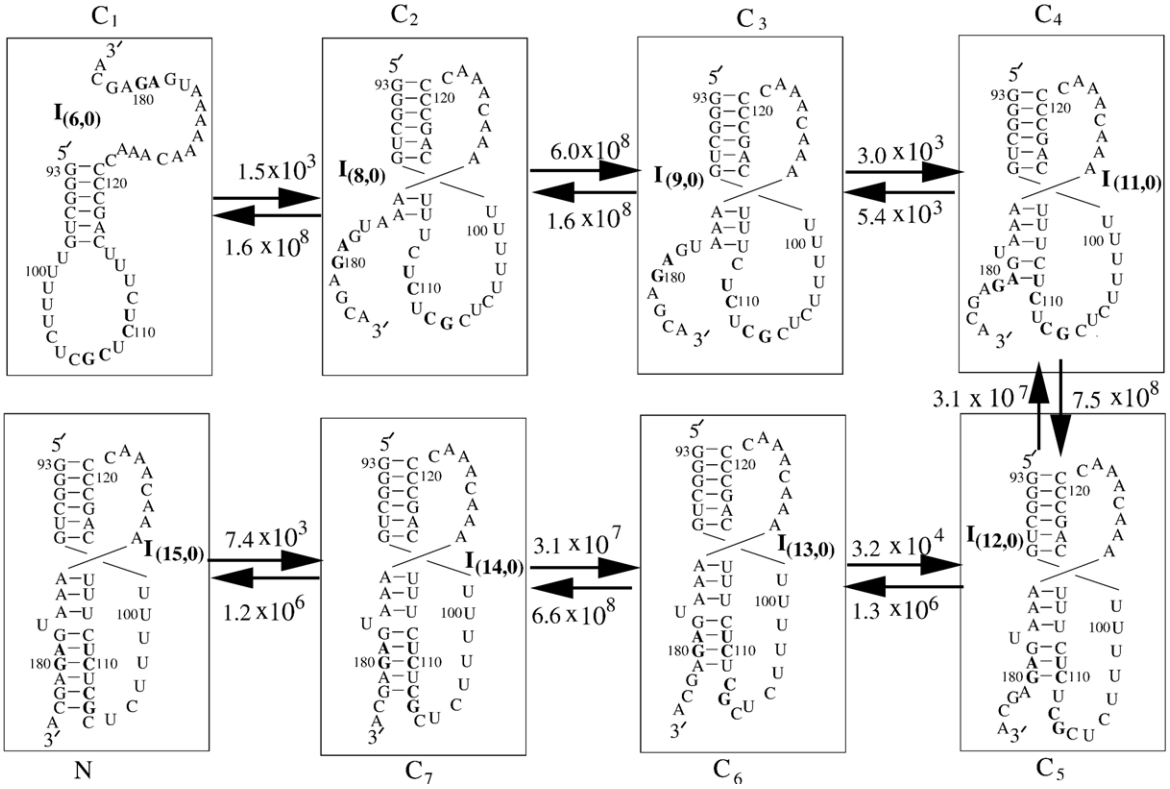

Figure 11. The detailed folding pathway from the hairpin $I_{(6,0)}$ to the pseudoknot $I_{(15,0)}$ for $PK_{179AG/110CU}$ at $T = 37$ °C. Also shown in the figure are the inter-cluster rate constants (in s^{-1}).

causing the loop (J2b/3)–stem (P3) tertiary interactions. As a result, for the A174 residue in P3, its pairing with a uridine residue (e.g., U99) in the J2b/3 loop would lead to the disruption of its pairing with the U115 residue in the P3 stem. Therefore, the wild-type hTR is predicted to have a different structure (without the A174:U115 base-pair in P3) from the NMR structure (with the A174:U115 base-pair), which does contain the bulged U177 residue.

The current form of the folding kinetics model developed in this study does not treat loop–stem tertiary interactions. The model gives the pseudoknot folding kinetics prior to the formation of the tertiary interactions. The model cannot predict the tertiary disruption of the U115:A174 base-pair following the formation of the pseudoknot (structure *N* in Figure 9(a)). Moreover, the present theory is based on a reduced atomic model for RNA conformations, namely, the virtual bond model (*Vfold*) of RNA.[23,24,32] In the *Vfold* model, a conformation is defined by the conformation of the nucleotide backbone and the base-pairing constraints. The model does not keep track of the detailed positions for each individual atoms in the sugar rings or in the chemical groups of the bases. Instead, the energy functions used in this model (implicitly) account for the interactions averaged over the ensemble of different atomic positions accessible to the given conformation.

For example, the model does not consider the particular position and the resultant interaction for the O_2' atom of A174 and the O_4' atom of A175, which were found to be in close proximity (and thus possibly repel each other) in the NMR experiment.[11]

Summary

The kinetic cluster approach enables a detailed analysis for RNA pseudoknot folding and unfolding kinetics, including rates, pathways, rate-limiting steps, activation energies, etc. Applications to PK5 and hTR pseudoknots reveal biphasic (two-step) folding kinetics: the chain is rapidly trapped in a misfolded 5′ hairpin state at $t \sim 10^{-3}$ s, followed by a detrapping transition from the misfolded hairpin to the native pseudoknot. The hairpin emerges as a kinetic intermediate because the hairpin → pseudoknot transition, which involves the breaking of the misfolded base-pairs and the formation of the pseudoknot loops, is slow.

In general, the hairpin → pseudoknot transition is the rate-limiting step of the overall folding process and the disruption of the misfolded base-pairs gives a positive activation energy.

For the hTR pseudoknot, the hairpin → pseudoknot mechanism may be essential for the telomerase activity and the function may be kinetically rather than thermodynamically controlled. For the experimentally studied mutation PK_{DC} and $PK_{\Delta U177}$, either the hairpin or the pseudoknot, respectively, is absent in the kinetic pathways, causing the reduction in telomerase activity for both PK_{DC} and $PK_{\Delta U177}$. For the hairpin-destabilizing mutation $PK_{179AG/110CU}$,

however, the hairpin is found to emerge as a kinetic intermediate in the folding process. Therefore, $PK_{179AG/110CU}$ preserves the telomerase activity.

The present model neglects the loop–stem tertiary interactions. For PK5 pseudoknot, the predicted pseudoknot stability without loop–stem interaction (4.2 kcal/mol) is close to the experimental result (4.3 kcal/mol (NMR) or 4.9 kcal/mol (optical)),[38] suggesting a weak loop–stem interaction. Therefore, the loop–stem tertiary interaction would not cause significant changes in the predicted folding kinetics. For the hTR pseudoknot, though the loop–stem tertiary interaction ($T_m \sim 40$ °C) is less stable than pseudoknot ($T_m \sim 60$ °C),[7,11] it can play significant structural and functional roles. Extensive RNA folding experiments have suggested that base-pairing patterns in the helix stems form first followed by the subsequent formation of the tertiary interactions.[55–58] Here we assume that the loop–stem tertiary interactions are formed in the late stage after the pseudoknot (in the absence of the loop–stem tertiary interaction) is folded. Based on such an assumption, the results predicted in the present study would provide insights into the kinetics, including the formation of the transient hairpin intermediates, the rate-limiting steps, the rates, and the pathways, prior to the formation of the tertiary interaction. However, as a caveat, we note that the present study assumes a 1 M NaCl solution condition and neglects the effect of the Mg^{2+} ions. Mg^{2+} can significantly stabilize the tertiary interactions[59–61] and thus, depending on the nucleotide sequence and structural distribution, may possibly alter the folding pathways. Further development of the pseudoknot folding kinetics model should include loop–stem tertiary interactions. In fact, with the use of the *Vfold* model,[23,32] which can give the chain entropy, it would be possible to obtain the loop–stem interaction energies from the thermodynamic experiments. Future development of the model by incorporating the tertiary interactions parameters and the Mg^{2+} effects into the model would allow us to treat the complete RNA pseudoknot folding kinetics (including tertiary structural folding).

Acknowledgements

We thank Dr Wenbing Zhang for valuable discussions about the kinetic model. We acknowledge grant support from NIH (GM063732 to S-J.C).

References

1. Draper, D. E. (1990). Pseudoknots and the control of protein synthesis. *Curr. Opin. Cell Biol.* **2**, 1099–1103.
2. Gesteland, R. F. & Atkins, J. F. (1996). Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.* **65**, 741–768.
3. Giedroc, D. P., Theimer, C. A. & Nixon, P. L. (2000). Structure, stability and function of RNA pseudo-knots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.* **298**, 167–185.
4. Plant, E. P., Jacobs, K. L. M., Harger, J. W., Meskauskas, A., Jacobs, J. L., Baxter, J. L. *et al.* (2003). The 9 Å solution: how mRNA pseudoknots promote efficient programmed −1 ribosomal frameshifting. *RNA*, **9**, 168–174.
5. Plant, E. P. & Dinman, J. D. (2005). Torsional restraint: a new twist on frameshifting pseudoknots. *Nucl. Acids Res.* **33**, 1825–1833.
6. Kim, Y. G., Su, L., Maas, S., O'Neill, A. & Rich, A. (1999). Specific mutations in a viral RNA pseudo-knot drastically change ribosomal frameshifting efficiency. *Proc. Natl. Acad. Sci. USA*, **96**, 14234–14239.
7. Cornish, P. V., Hennig, M. & Giedroc, D. P. (2005). A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated-1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. USA*, **102**, 12694–12699.
8. Blackburn, E. H. (1998). Telomerase RNA Structure and Function. In *RNA Structure and Function* (Simons, R. W. & Grunberg-Manago, M., eds), pp. 669–693, Cold Spring Harbor Lab. Press, New York.
9. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H. & James, T. L. (2002). A molecular switch underlies a human telomerase disease. *Proc. Natl. Acad. Sci. USA*, **99**, 16998–17003.
10. Theimer, C. A., Finger, L. D., 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.
11. Theimer, C. A., Blois, C. A. & Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell*, **17**, 671–682.
12. Theimer, C. A. & Feigon, J. (2006). Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16**, 307–318.
13. Chen, J. L. & Greider, C. W. (2005). Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc. Natl. Acad. Sci. USA*, **102**, 8080–8085.
14. Chen, J. L. & Greider, C. W. (2004). Telomerase RNA structure and function: implications for dyskeratosis congenita. *Trends Biochem. Sci.* **29**, 183–192.
15. Marrone, A., Walne, A. & Dokal, I. (2005). Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr. Opin. Genet. Dev.* **15**, 249–257.
16. Batey, R. T., Rambo, R. P. & Doudna, J. A. (1999). Tertiary motifs in RNA structure and folding. *Angew. Chem., Int. Ed. Engl.* **38**, 2326–2343.
17. Tanner, N. K., Schaff, S., Thill, G., Petitkoskas, E., Crainden Petitkoskas, E., Craindenoyelle, A. M. & Westhof, E. (1994). A 3-dimensional model of hepatitis-delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* **4**, 488–498.
18. Ke, A. L., Zhou, K. H., Ding, F., Cate, J. H. D. & Doudna, J. A. (2004). A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, **429**, 201–205.
19. Mathews, D. H. & Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **16**, 270–278.
20. McCaskill, J. S. (1990). The equilibrium partition-function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
21. Chen, S. J. & Dill, K. A. (2000). RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA*, **97**, 646–651.
22. Ding, Y. (2006). Statistical and Bayesian approaches to RNA secondary structure prediction. *RNA*, **12**, 323–331.

23. Cao, S. & Chen, S. J. (2005). Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*, **11**, 1884–1897.
24. Cao, S. & Chen, S. J. (2006). Free energy landscapes of RNA/RNA complexes: with applications to snRNA complexes in spliceosomes. *J. Mol. Biol.* **357**, 292–312.
25. Orland, H. & Zee, A. (2002). RNA folding and large N matrix theory. *Nucl. Phys., B*, **620**, 456–476.
26. Rivas, E. & Eddy, S. R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **285**, 2053–2068.
27. Lyngsø, R. B. & Pedersen, C. N. S. (2000). RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.* **7**, 409–427.
28. Dirks, R. M. & Pierce, N. A. (2003). A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.* **24**, 1664–1677.
29. Ren, J., Rastegari, B., Condon, A. & Hoos, H. H. (2005). HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA*, **11**, 1494–1504.
30. Lucas, A. & Dill, K. A. (2003). Statistical mechanics of pseudoknot polymers. *J. Chem. Phys.* **119**, 2414–2421.
31. Kopeikin, Z. & Chen, S. J. (2006). Folding thermodynamics of pseudoknotted chain conformations. *J. Chem. Phys.* **124**, 154903.
32. Cao, S. & Chen, S. J. (2006). Predicting RNA pseudoknot folding thermodynamics. *Nucl. Acids Res.* **34**, 2634–2652.
33. Flamm, C., Fontana, W., Hofacker, I. L. & Schuster, P. (2000). RNA folding at elementary step resolution. *RNA*, **6**, 325–338.
34. Zhang, W. & Chen, S. J. (2002). RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci. USA*, **99**, 1931–1936.
35. Yingling, Y. G. & Shapiro, B. A. (2005). Dynamic behavior of the telomerase RNA hairpin structure and its relationship to dyskeratosis congenita. *J. Mol. Biol.* **348**, 27–42.
36. Isambert, H. & Siggia, E. D. (2000). Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. USA*, **97**, 6515–6520.
37. Puglisi, J. D., Wyatt, J. R. & Tinoco, I., Jr (1990). Conformation of an RNA pseudoknot. *J. Mol. Biol.* **214**, 437–453.
38. Wyatt, J. R., Puglisi, J. D. & Tinoco, I., Jr (1990). RNA pseudoknots—stability and loop size requirements. *J. Mol. Biol.* **214**, 455–470.
39. Serra, M. J. & Turner, D. H. (1995). Predicting thermodynamic properties of RNA. *Methods Enzymol.* **259**, 242–261.
40. Kierzek, R., Burkard, M. E. & Turner, D. H. (1999). Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, **38**, 14214–14223.
41. Zhang, W. B. & Chen, S. J. (2006). Exploring the complex folding kinetics of RNA hairpins. 1. General folding kinetics analysis. *Biophys. J.* **90**, 765–777.
42. Zhang, W. B. & Chen, S. J. (2003). Analyzing the biopolymer folding rates and pathways using kinetic cluster method. *J. Chem. Phys.* **119**, 8716–8729.
43. Chan, H. S. & Dill, K. A. (1994). Transition-states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* **100**, 9238–9257.
44. Pörschke, D., Uhlenbeck, O. C. & Martin, F. H. (1973). Thermodynamics and kinetics of the helix–coil transition of oligomers containing GC base-pairs. *Biopolymers*, **12**, 1313–1335.
45. Munõz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature*, **390**, 196–199.
46. Bonnet, G., Krichevsky, O. & Libchaber, A. (1998). Kinetics of conformational fluctuations in DNA hairpin–loops. *Proc. Natl. Acad. Sci. USA*, **95**, 8602–8606.
47. Wallace, M. I., Ying, L. M., Balasubramanian, S. & Klenerman, D. (2001). Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc. Natl. Acad. Sci. USA*, **98**, 5584–5589.
48. Wang, X. J. & Nau, W. M. (2004). Kinetics of end-to-end collision in short single-stranded nucleic acids. *J. Am. Chem. Soc.* **126**, 808–813.
49. Onuchic, J. N., LutheySchulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600.
50. Ansari, A., Kuznetsov, S. V. & Shen, Y. (2001). Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl. Acad. Sci. USA*, **98**, 7771–7776.
51. Dube, S. K., Marcker, K. A., Clark, B. F. & Cory, S. (1968). Nucleotide sequence of *N*-formyl-methionyl-transfer RNA. *Nature*, **218**, 232–233.
52. Cole, P. E. & Crothers, D. M. (1972). Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, **11**, 4368–4374.
53. Chen, J. L., Blasco, M. A. & Greider, C. W. (2000). Secondary structure of vertebrate telomerase RNA. *Cell*, **100**, 503–514.
54. Yingling, Y. G. & Shapiro, B. A. (2006). The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation. *J. Mol. Graph. Model.* **25**, 261–274.
55. Zarrinkar, P. P. & Williamson, J. R. (1994). Kinetic intermediates in RNA folding. *Science*, **265**, 918–924.
56. Rangan, P., Masquida, B., Westhof, E. & Woodson, S. A. (2003). Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. USA*, **100**, 1574–1579.
57. Woodson, S. A. (2005). Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324–330.
58. Kwok, L. W., Shcherbakova, I., Lamb, J. S., Park, H. Y., Andresen, K., Smith, H. *et al.* (2006). Concordant exploration of the kinetics of RNA folding from global and local perspectives. *J. Mol. Biol.* **355**, 282–293.
59. Draper, D. E. (2004). A guide to ions and RNA structure. *RNA*, **10**, 335–343.
60. Draper, D. E., Grilley, D. & Soto, A. M. (2005). Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 221–243.
61. Grilley, D., Soto, A. M. & Draper, D. E. (2006). Mg^{2+}–RNA interaction free energies and their relationship to the folding of RNA tertiary structures. *Proc. Natl. Acad. Sci. USA*, **103**, 14003–14008.
62. Olson, W. K. (1980). Configurational statistics of polynucleotide chains: an updated virtual bond model to treat effects of base stacking. *Macromolecules*, **13**, 721–728.

Edited by D. E. Draper

(Received 3 November 2006; received in revised form 22 December 2006; accepted 2 January 2007)
Available online 9 January 2007