A New Computational Approach for Mechanical Folding Kinetics of RNA Hairpins

Song Cao and Shi-Jie Chen*
Department of Physics and Astronomy and Department of Biochemistry, University of Missouri, Columbia, Missouri

ABSTRACT Based on an ensemble of kinetically accessible conformations, we propose a new analytical model for RNA folding kinetics. The model gives populational kinetics, kinetic rates, transition states, and pathways from the rate matrix. Applications of the new kinetic model to mechanical folding of RNA hairpins such as trans-activation-responsive RNA reveal distinct kinetic behaviors in different force regimes, from zero force to forces much stronger than the critical force for the folding-unfolding transition. In the absence of force or a low force, folding can be initiated (nucleated) at any position by forming the first base stack and there exist many pathways for the folding process. In contrast, for a higher force, the folding/unfolding would predominantly proceed along a single zipping/unzipping pathway. Studies for different hairpin-forming sequences indicate that depending on the nucleotide sequence, a kinetic intermediate can emerge in the low force regime but disappear in high force regime, and a new kinetic intermediate, which is absent in the low and high force regimes, can emerge in the medium force range. Variations of the force lead to changes in folding cooperativity and rate-limiting steps. The predicted network of pathways for trans-activation-responsive RNA suggests two parallel dominant pathways. The rate-limiting folding steps (at $f = 8$ pN) are the formation of specific basepairs that are 2–4 basepairs away from the loop. At a higher force ($f = 11$ pN), the folding rate is controlled by the formation of the bulge loop. The predicted rates and transition states are in good agreement with the experimental data for a broad force regime.

INTRODUCTION

Elucidation of RNA folding mechanism at the level of both secondary and tertiary structures is essential to understanding RNA functions in transcription, splicing, and translation. RNA hairpins are important for biological functions, partly because they can be uniquely recognized by different proteins (1). In a cell, RNAs often undertake the force exerted by its environments, e.g., the messenger RNA can be pulled into the decoding site so that ribosome can read the message (2,3). Moreover, many RNA functions are kinetically controlled. Therefore, modeling RNA hairpin folding kinetics under force is significant for quantitative predictions of functions and rational design of RNAs for therapeutic applications.

Recently, an increasing amount of theoretical and experimental studies for force-free folding (4–7) led to the conclusion that even for simple hairpins, the free energy landscape can be quite rugged and folding can involve misfolded intermediates (4–15). These studies on RNA hairpin folding energy landscapes suggested several different mechanisms for the rate-limiting steps, from loop closure, disruption of misfolded states, or single-stranded base stacks (for loop formation), to slow formation of specific base stacks in the helix stem. The kinetic rates show strong temperature and sequence-dependence. At high temperatures or a temperature around the folding-unfolding transition, folding can be two-state and have a positive or a negative activation barrier; however, for lower temperatures, folding can become noncooperative and involve a positive apparent activation barrier, presumably corresponding to the energy cost to break the misfolded states. Recently, a theory based on the kinetic cluster method systematically explored a broad range of kinetic behaviors for RNA hairpin folding for different temperatures and different helix and loop sequence contexts and lengths (15).

Recent experiments on force-induced RNA folding probed by atomic force microscopes, laser tweezers, and magnetic tweezers (16–19) greatly expanded the scope of RNA folding (20–29). Mechanical folding in these experiments provided a highly effective tool for probing the mechanical properties of RNA folding, mapping the free energy landscape, and controlling the folding-unfolding trajectories. From the force-extension curve and the time series of the extension, the mechanical folding experiments reveal kinetic information for the conformational switches and the presence or absence of kinetic intermediates. For example, the time-dependence of the end-to-end distance for P5ab hairpin under an external force suggests that the folding-unfolding transition is two-state, without detectable intermediates (16).

These remarkable experiments have greatly inspired theoretical studies of force-induced unfolding (30–40). Theoretical studies found that the force-extension curve of a single-stranded RNA could be well modeled as an elastic, freely jointed chain (31). Moreover, theoretical predictions for unfolding along one pathway (unzipping from the end of helix stem) yielded good agreements with experimental data (41). RNA can fold/unfold along many different pathways, and the folding kinetic behavior can strongly depend on the force range. Therefore, studies based on a specific force range used in experiments and theories may not reveal the full landscape

Submitted November 5, 2008, and accepted for publication February 20, 2009.

*Correspondence: chenshi@missouri.edu

Editor: Angel E. Garcia.

0006-3495/09/05/4024/11 $2.00

doi: 10.1016/j.bpj.2009.02.044

of the force-dependence of the folding kinetics. It is thus desirable to investigate the folding kinetics, including the folding rate, transition state, kinetic intermediates, and folding pathways, for a broad force range.

The force range probed experimentally and theoretically was limited to the narrow range around the point at which the folding/unfolding transition takes place. In this article, based on a new kinetic model, we discuss the effect of force on the folding/unfolding kinetics for several sequences in the full range of the force from zero to the point at which the hairpin is fully unfolded. We aim to understand the general principles for the effect of force on the folding/unfolding kinetics. We will first develop a new kinetic theory that allows us to effectively account for the kinetically accessible conformational ensemble. We will then apply the new theory to study the mechanical folding-unfolding of RNA hairpins under external pulling force. We will compare our theoretical prediction with the experimental data. We will also combine the free energy landscape analysis and the kinetic master equation to investigate the force-dependence of the folding rate and the folding pathways and rate-limiting steps for different sequences. Because the model is mainly analytical and deterministic, it enables stable, deterministic predictions for the long time dynamics and detailed analysis for the folding mechanisms. Furthermore, the new kinetic model developed here, which is validated through experimental tests, can effectively reduce a huge conformational ensemble down to a manageable size. This new model may be potentially useful for future theoretical studies regarding more-complex RNAs, beyond hairpins.

KINETIC ANALYSIS

Based on the virtual bond representation of polynucleotide conformations, we recently developed a new model (i.e., V-fold) for RNA conformational entropy and folding free energy landscape. Experimental tests for the thermally induced equilibrium RNA folding thermodynamics indicate that the model is reliable (42,43). We first test the validity of the new conformational model for mechanical folding-unfolding through theory-experiment comparisons for the force-extension curves in the equilibrium folding process. We show the thermodynamic analysis for the constant force ensemble and constant distance ensemble in the Supporting Material. Our theoretical predictions agree with the experimental results by Li et al. (24). In what follows, we will focus on the kinetics of the mechanical folding of RNA hairpins.

A new kinetic model

Due to the large number of possible states for a long chain, the master equation (7) (see Supporting Material) approach is limited to short sequences. The kinetic cluster theory (15,44,45) (see Supporting Material), based on the classification of conformations into reduced states, can effectively reduce the size of the conformational ensemble. Our kinetic cluster approach is useful for kinetic systems with a manageable number of clusters (~6000). However, the kinetic cluster method would become highly convoluted for larger RNAs, which can have a much larger number of clusters. Thus, it is desirable to develop a new kinetic model which, in combination with the kinetic cluster method, can treat the kinetics efficiently for large RNAs. Such a model may have the advantage over the simulational methods (46–48), which are often limited by stochastic conformational sampling.

We note that a vast number of conformational states may not be significantly sampled in the folding process. These states are usually slow to form (due to large entropic loss) and fast to disrupt (with small enthalpic cost). As a result, the stability of these states $\Delta G = \Delta H - T\,\Delta S$ is usually low (high free energy ΔG). Based on the above observation, we developed the following new method for generating the conformational ensemble for the kinetic analysis.

1. We select the first Ω_c low-free energy states whose (force-free) folding stability ΔG is lower than a threshold ΔG_c (Fig. 1 *a*). We treat each such state as a parent state. Base-pairs contained in a parent are called parent basepairs. Note that a parent state can be either nativelike or misfolded.
2. For each parent state i (i = 1, 2, 3,…, and Ω_c), we enumerate all the daughter states that stem from the parent state. A daughter state contains 0, 1, 2,…., parent basepairs and no non-parent basepairs (see Fig. 1, *b* and *c*). Moreover, a daughter state can form all the possible loops that satisfy the following conditions: $x'_1 = x_1$ and $x'_2 = x_2$; $x'_3 \le x_3$ and $y'_3 \ge y_3$; and $y'_2 = y_2$ and $y'_1 = y_1$.
3. All the parents and their daughters form a reduced conformational ensemble for the kinetic analysis. We construct the rate matrix for this reduced conformational ensemble. The size of the reduced conformational ensemble is much smaller and grows much more slowly with the chain

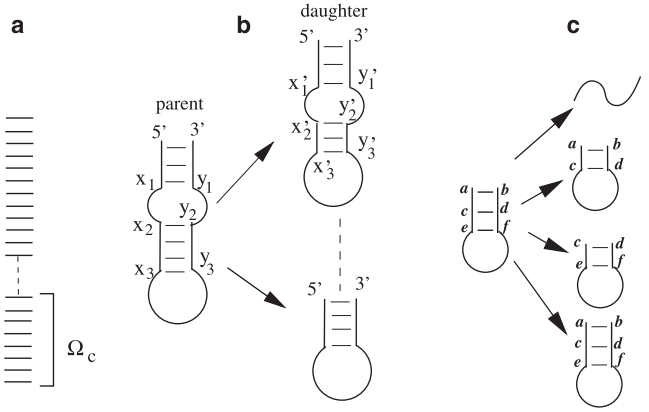

FIGURE 1 (*a*) The selected Ω_c conformations based on the thermal stability. (*b*) A parent and the possible daughter states derived from the parent. (*c*) A parent with two base stacks. The four daughter states include a coil state, two states with only one base stack, and one state with two base stacks.

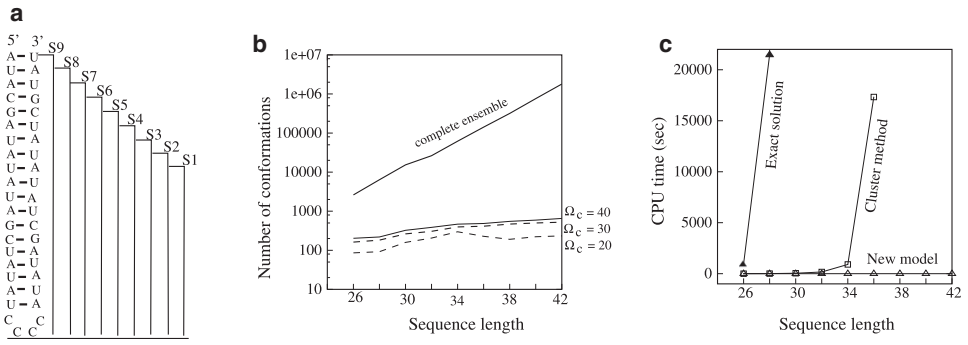

FIGURE 2 (*a*) S1–S9 are the nine sequences with different stem lengths and hence different sequence lengths. These sequences are used to compute the data for panel *b*. (*b*) The number of conformations from the complete conformational ensemble model and the new kinetic model based on the number of parent states $\Omega_c = 20$, 30, and 40. We find the number of conformations from the complete conformational ensemble model grows exponentially as the chain length increases. However, the number of conformations from the new kinetic model increases much more slowly with the chain length. (*c*) The computational time for the exact master equation (*solid triangle*), kinetic cluster method (*open rectangle*), and the new kinetic method ($\Omega_c = 20$) (*open triangle*). The new kinetic method can significantly improve the computational speed. The computational time for the nine sequences is <1 s. In contrast, both the exact master equation and the cluster method are computationally expensive and can only treat short RNA sequences.

length than the original complete ensemble (see Fig. 2). Depending on the sequence length, the number of conformations can be reduced by a factor of 10 for a 26-nt chain and $>10^3$ for a 42-nt chain. The smaller size of the conformational ensemble enables viable and efficient computations with the master equation and the kinetic cluster analysis.

In practice, to choose the number of parent conformations for a given RNA sequence, we perform a series of tests from a small Ω_c to larger Ω_c values until the results of the relaxation rates (eigenvalues of the rate matrix) converges. For the sequences to be investigated in this study (sequence $\leq$120 nt), we find that $\Omega_c = 20$ is sufficient for a robust prediction of the folding kinetics. The tested sequences are HP1 and HP2 in Fig. 3, *a* and *b*, HP3 and HP4 in Fig. S7, *a* and *b*, trans-activation-responsive (TAR) RNA in Fig. 6 *a*, and S13 in Fig. S5 *b*.

Force-dependence of the rate and the rate-limiting step

Because hydrogen-bonding and base-stacking interactions are short-ranged, the transition from the basepaired state to the unpaired state may occur over a small displacement (~0.1 nm) (41). In an earlier article, Cocco et al. (41) estimated that the mechanical work done by a pulling force f is (0.1 nm) $f \sim 0.3\ k_BT$ for a typical force of 15 pN at room temperature. The work is much smaller than the base-stacking enthalpy parameter ΔH, which is $\geq 6\ k_BT$ (49). Therefore, for the disruption of a basepair, we neglect the perturbation of the force on the kinetic barrier. In contrast, for the formation of a basepair, two originally separated (unpaired) nucleotides are juxtaposed into close proximity from a larger separation; thus, the force causes a large excess kinetic barrier. The work done by the force can be estimated as $2g_s(f)$, where $g_s(f)$ is the free energy per nucleotide under external force f per nucleotide (see Eq. S8 and Fig. S1 *b* in Supporting Material) and the factor 2 is the number of the nucleotides involved in the formation of the base pair. Therefore, we obtain the following rate constants for the formation (k_+) and disruption (k_-) of a basepair (or a base stack):

$$k_+(f) = k_+(0)e^{-2g_s(f)/k_BT} = k_0 e^{-(T\Delta S + 2g_s(f))/k_BT}, \quad (1)$$

$$k_-(f) = k_0\, e^{-\Delta H/k_BT}. \quad (2)$$

The kinetic barrier for the formation of a basepair comes from two contributions: $T\Delta S$ from the entropy decrease and $2g_s(f)$ from the force. The barrier to break a basepair comes from the ΔH of the enthalpy increase of the base stack. Therefore, the rate-limiting steps for folding would

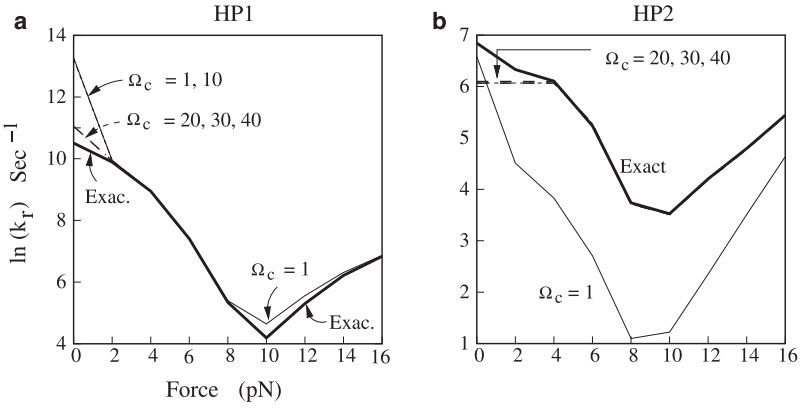

FIGURE 3 The force dependence of the relaxation rate for (*a*) HP1 (Fig. S3 *a*) and (*b*) HP2 (Fig. S3 *b*) at $T = 25°C$. We compare the results from the complete ensemble model and from the new kinetic model. We test the dependence of the predicted rate constant on the parameter Ω_c (the number of parent states) used in the new kinetic model. For HP1, we find that the folding rate is the same for $\Omega_c = 1$ and 10 at force $\leq$ 2 pN. For force $\geq$ 2 pN, the predicted rate constant is the same for $\Omega_c = 10$, 20, 30, and 40.

be 1), the formation of the native base stacks of the largest $T\Delta S + 2g_s(f)$; and/or 2), the disruption of nonnative stacks of the largest ΔH. In contrast, the rate-limiting steps for unfolding would be the disruption of the native base stacks of the largest ΔH. To estimate how strong the effect of force is on the rate constant, we calculate the force-induced barrier $2g_s(f)$. We found that for temperature $T = 25°C$ and force $f = 12$ pN, $g_s(f) = 0.64$ kcal/mol, which is small compared to $T\Delta S$ and ΔH of a base stack. Therefore, the effect of the force on the slowest folding step is weak.

Although the force may not cause a significant change to the rate constant for a single kinetic move (Eqs. 1 and 2), its effect on a structure containing multiple base stacks can be significant due to the base-to-base accumulation. From Eq. S12, the overall rate is determined not only by the rate of individual kinetic moves, but also by the populational distribution of the states. In the kinetic cluster framework, the populational distribution is related to the stabilities of the conformations in the respective local preequilibrated cluster. The pulling force can significantly change the populational distribution of different states, which often contain multiple base stacks. As a result, the external force can cause notable changes to the overall folding and unfolding rate and the partitioning of the pathways.

Force dependence of the kinetics

We first study the folding kinetics for two short hairpin-forming sequences. These sequences are short. Therefore, we can exactly solve the kinetics based on the complete conformational ensemble. We will also apply our new kinetic model, which is based on the reduced conformational ensemble, to solve the kinetics. Our motivation here is twofold: 1), to evaluate the effect of force on the misfolded intermediates and the transition states in different force regimes; and 2), to validate the new kinetic model for hairpin folding kinetics. Because the temperature dependence of folding kinetics has been thoroughly studied (36,50), below we mainly focus on the force dependence of folding and unfolding kinetics at a fixed temperature $T = 25°C$ (unless explicitly stated otherwise).

We will then apply the model to study the folding kinetics of TAR RNA. The complete conformational ensemble of TAR RNA (52-nt) is too large to be treated by the master equation method or the kinetic cluster method. Our control tests suggested that the maximum number of conformations that we can treat using the kinetic cluster method is ~10^6, corresponding to chain length ~40 nt for a generic sequence (Fig. 2, *a* and *b*). In addition, for such large number of conformations, the kinetic cluster method yields ~6000 clusters for the system. It takes ~7 h to obtain the eigenvectors and eigenvalues for the rate matrix on a Dell EM65T cluster computer system with one Intel Xeon 5150 (2.66 GHz) processor. Fig. 2 *c* shows the computational time for different kinetic models. The new kinetic model significantly improves the computational efficiency. For example, the new model takes <1 s to compute the folding rate for a 42-nt sequence. Further tests suggest that the new method can efficiently handle much larger RNA hairpins (>100 nt) (see Fig. S4 and Fig. S5, *a* and *b*). In contrast, both the exact master equation method based on the complete conformational ensemble and the kinetic cluster method are computationally expensive and can only treat much shorter sequences. To treat a long sequence such as TAR RNA, the new model developed here is essential.

UAUAGCUAUAUCCCCAUAUAGCUAUA

For this sequence (labeled as HP1), the force-free native structure is a hairpin with loop-length $L_{loop} = 4$ (see Fig. S3 *a*). With the external force, the thermodynamic model based on the complete conformational ensemble predicts that at temperature $T = 25°C$, the hairpin unfolds at critical force $f_c = 10.0$ pN for a constant force ensemble. At $f = f_c$, the native state is almost unzipped completely (data not shown).

Analysis based on the exact complete conformational ensemble model. For the 26-nt sequence, exhaustive enumeration of all the possible states (described by base stacks) gives 2617 conformational states. For the force range $0 \le f \le 16$ pN (see Fig. 3 *a*), the eigenvalue spectrum of the rate matrix for the complete conformational ensemble has a gap between the first (λ_1) and the second (λ_2) nonzero eigenvalues (rates). Therefore, the slowest mode (λ_1) in Eq. S10 controls the folding rate and the folding time is $\sim t_{fold} \sim 1/\lambda_1$.

Formation of the misfolded intermediates in the low-force limit. Without the external force, the populational kinetics of the native state follows $P_N(t) \simeq P_N^{eq} - 0.67\ e^{-\lambda_1 t}$, with $\lambda_1 = 3.6 \times 10^4\ s^{-1}$ and the equilibrium fractional population of the native hairpin $P_N^{eq} = 0.83$. Among all the native base stacks, we find that 5GC6-21GC22 has the largest $|\Delta S|$, thus the formation of this native base stack is an on-pathway rate-limiting step. Among all the nonnative base stacks, 7UA8-25UA26 has the largest $|\Delta H|$, thus, detrapping from the misfolded base stack 7UA8-25UA26 (see I_2 in Fig. S6 *a*) is an off-pathway rate-limiting step. In summary, folding of the hairpin is biphasic:

1. Rapid formation of the misfolded state I_2. The large entropic barrier for the on-pathway rate-limiting step (formation of the native stack 5GC6-21GC22) and, thus, the slow formation of the native structure, causes the formation of the misfolded state.
2. Slow detrapping from the misfolded state I_2. Detrapping from I_2 is the major rate-limiting step for the overall folding process. As shown in Fig. 4 *a*, the populational kinetics solved from the master equation for the 2617 states shows that the misfolded state I_2 indeed emerges as an intermediate in the absence of the external force.

Force-induced destabilization of the misfolded intermediates. Increasing the force can lower the barrier for breaking

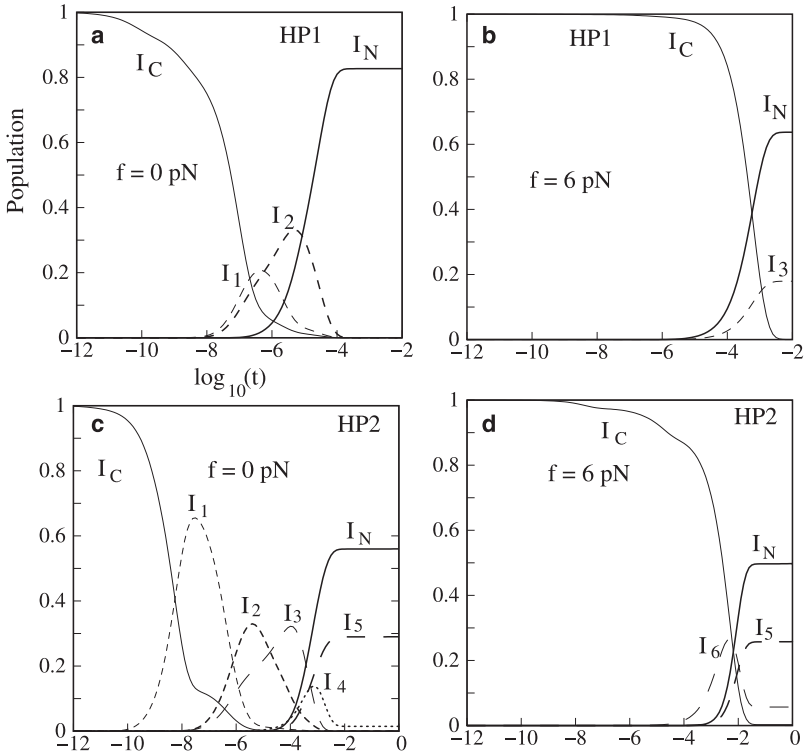

FIGURE 4 The populational kinetics for HP1 (Fig. S3 *a*) and HP2 (Fig. S3 *b*) at different forces.

the misfolded state. For detrapping from the misfolded states, though the rate to break a (misfolded) base stack ($\sim e^{-\Delta H/k_BT}$) is assumed to be force-independent, the probability for the chain to take the folding pathway through detrapping is force-sensitive due to the force-dependent stability of the misfolded states. A large force would destabilize the misfolded intermediates and cause the chain to have a smaller probability to take the folding route through the intermediates. For example, in the absence of the force, intermediates I_1 and I_2 appear as a transient kinetic intermediate. For force $f = 6$ pN, the intermediates I_1 and I_2 disappear (see Fig. 4 *b*). The folding pathways through I_1 and I_2 become kinetically silent (probability $\simeq 0\%$). A nativelike on-pathway intermediate I_3 appears. Therefore, through changing the stability of the (misfolded) kinetically trapped state, the force can significantly alter the partitioning of the kinetic pathways, resulting in the change of the kinetic cooperativity of folding. In this example, an increase of the force would turn a noncooperative (non-single-exponential) folding reaction into a cooperative (single-exponential) kinetics.

Analysis based on the new kinetic model with reduced conformational ensemble

Validity of the new kinetics model. Using the new kinetic model, we can greatly reduce the number of conformations (and hence the size of the rate matrix). For the 26-nt hairpin sequence, the total numbers of conformations in the reduced ensemble are 50, 56, 86, 162, and 203 for the number of parent conformations $\Omega_c = 1$, 10, 20, 30, and 40, respectively. From Fig. 3 *c*, we find that the predicted folding rates for different Ω_c values are convergent for $\Omega_c > 20$. Comparisons with the results from the complete conformational ensemble show excellent agreement except for the low-force limit $f \leq 2$ pN, where the new kinetic model underestimates the folding rate. This is because at low force, more intermediates, such as I_1 and I_2 in Fig. 4 *a*, can form. These states may be out of the first Ω_c states and thus neglected in the reduced conformational ensemble. Since these intermediates can contribute to the folding process, neglecting their contribution leads to underestimation of the rate. Under a large pulling force f, I_1 and I_2 become unstable compared to the coil state: $\Delta G \sim 7g_s(f)$ for I_1 with a 7-nt single-stranded RNA tail and $6g_s(f)$ for I_2 with a 6-nt tail compared to $26g_s(f)$ for the coil state, respectively (see Fig. S1 *b* for $g_s(f)$). I_1 and I_2 disappear at large forces, and neglecting them will not cause notable error in the prediction of the folding rates. Therefore, the new kinetic model works better for the forces that are not too low. In contrast, the increase of the force stabilizes state I_3, which has two unpaired nucleotides, $\Delta G \sim 2g_s(f)$. Therefore, I_3 instead of I_1 or I_2 emerges as a subpopulated states for $f = 6$ pN.

Folding under folding condition $f < f_c$. We first investigate the folding kinetics under folding condition $f = 6$ pN $< f_c$ ($= 10$ pN). To obtain the detailed information about the folding pathways, rates, and rate-limiting steps, we perform the kinetic cluster analysis for the ensemble of 86 conformations (based on the $\Omega_c = 20$ parent states). We first identify the slow-forming native stacks and slow-breaking nonnative base stacks. We find five rate-limiting stacks 1UA2-

25UA26, 3UA4-23UA24, 5GC6-21GC22, 7UA8-19UA20, and 9UA10-17UA18. They are slow-forming native stacks. These stacks are slow to form due to the large entropic decrease accompanying the formation of the base stacks (49). Using the five stacks, we can classify the 86 conformations (with $\Omega_c = 20$ parent conformations) into 16 kinetic clusters.

The 16×16 rate matrix (for the 16 kinetic clusters) gives the folding rate of 1813 s^{-1}, which is close to the folding rate 1639 s^{-1} predicted from the exact master equation based on the complete conformational ensemble. In addition, the fast intercluster transitions give the dominant folding pathway (see Fig. 5 *a*). Furthermore, applying Eq. S15 to the dominant folding pathway in Fig. 5 *a* gives a folding rate of 1711 s^{-1}, which again is close to the rate 1813 s^{-1}. From Fig. 5 *a*, folding is rate-limited by the transition from cluster C_1 to C_2. The dominant pathway for $C_1 \rightarrow C_2$ is from conformation $I_{(2,\ 0)}$ in cluster C_1 to $I_{(3,\ 0)}$ in cluster C_2 through the slow formation of the base stack 9UA10-17UA18. We can estimate the rate from Eq. S12 as $p_{I_{(2,0)}} k_{I_{(2,0)} \rightarrow I_{(3,0)}} = (6.4 \times 10^{-3}\%)(2.9 \times 10^{7} s^{-1}) = 1856\ s^{-1}$. The slow rate is caused by the very low fractional population of state $I_{(2,\ 0)}$ in cluster C_1. This rate is much slower than the formation of a stack (in $I_{(2,\ 0)}$) from a fully unfolded state because the fully unfolded state has a large fractional population in cluster C_1.

Unfolding under folding condition $f < f_c$. Under the unfolding condition with a higher force $f = 14$ pN, the 16×16 rate matrix for the kinetic clusters gives nearly the same unfolding rate (504 s^{-1}) as that from the exact master equation(500 s^{-1}). On the predicted dominant pathway (Fig. 5 *b*) at $f = 14$ pN, the rate-limiting step is the unfolding process from cluster U_3 to U_4, which involves disruption of the 5GC6-21GC22 native stack. The GC-GC stack has the largest enthalpic parameter (9.4 kcal/mol) (49). The folding rate from cluster U_3 to U_4 is 1000 s^{-1} from Eq. S12, which is roughly on the same scale as the rate 504 s^{-1} solved from the exact master equation. The slow $U_3 \rightarrow U_4$ rate is not caused by the fractional population of $I_{(7,\ 0)}$ (32%). Instead, it is caused by the slow $I_{(7,\ 0)} \rightarrow I_{(6,\ 0)}$ rate (2967 s^{-1}) for the breaking of the 5GC6-21GC22 base stack.

Bending of the rate-force curve in the low-force regime. The rate-force dependence in Fig. 3 *a* shows bending of the curve at the low force region. The bending is a result of two effects. First, at low force, misfolded intermediates would be stabilized and emerge in the folding process (Fig. 4 *a*). The formation and detrapping of these misfolded intermediates would cause the slowdown of the folding kinetics. Second, the force-induced barrier is related to the elastic free energy $g_s(f)$. The nonlinearity of the $g_s(f)$ curve at the low force region (Fig. S1 *b*) would contribute to the low-force bending of the rate-force curve in Fig. 3 *a*.

CGCUCAAAUGUCUAAAAGACAGAGCG

We next study the effect of force on the folding of a more complex hairpin-hairpin with a bulge loop (see HP2 in Fig. S3 *b*). From Fig. 3 *b*, we find that 1), the folding rate predicted from the new kinetic model converges for $\Omega_c \geq 20$; and 2), the results from the new kinetic model and from the original, complete ensemble model show excellent agreement for forces not too weak ($f \geq 4$ pN). For weak forces $f \leq 4$ pN, the new kinetic model underestimates the folding rate.

The force-free populational kinetics (Fig. 4 *c*) shows the formation of five intermediates, including four misfolded states (I_1, I_2, I_3, and I_4); see Fig. S6 *b* for the structures. These states have low stabilities and thus are not included in the reduced conformational ensemble. We find that these misfolded states I_1 and I_2 are still populated at low force $f = 2$ pN. For low forces, the folding can be complex due to the formation of the misfolded states and multiple folding pathways. The conclusion is consistent with the Monte Carlo simulation by Hyeon and Thirumalai (51).

For a large force $f \geq 4$ pN, these misfolded states become kinetically inaccessible. For example, for the misfolded states I_1, because the misfolded state (kinetic trap) involves the formation of a 7-nt tail in the 3′ end of the helix, its relative stability compared to the coil state would be greatly

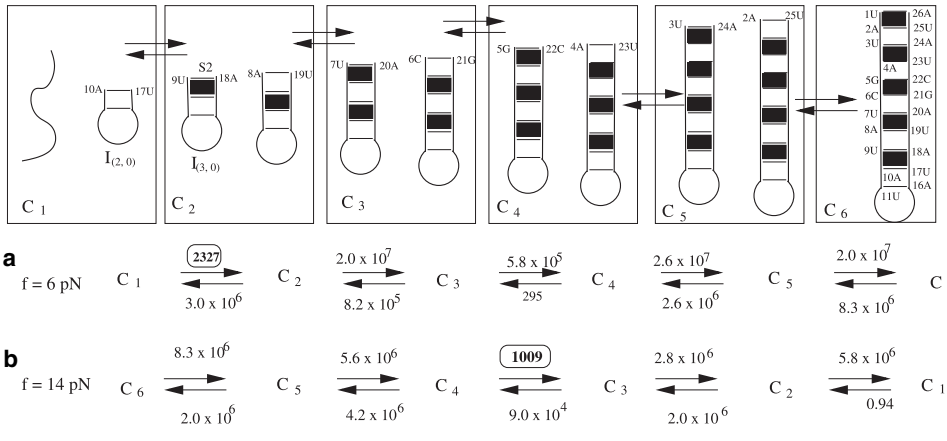

FIGURE 5 (*a*) The dominant folding pathway and the intercluster transition rates for the hairpin in Fig. 3 *a* at $f = 6$ pN and $T = 25°C$. (*b*) The unfolding pathway at $f = 14$ pN and $T = 25°C$. We find the rate-limiting step for folding is the formation of the native base stack 9UA10-17UA18 ($C_1 \rightarrow C_2$) due to the large entropic barrier, while the rate-limiting step for the unfolding is the disruption of the nonnative base stack 5GC6-21GC22 ($C_4 \rightarrow C_3$) due to the large enthalpic cost for breaking the base stack. $I_{(N,\ NN)}$ denotes a conformation with N native basepairs and NN nonnative basepairs. The figure also shows the intercluster transition rates in s^{-1}.

decreased as the force f is increased. For $f = 6$ pN, there exists only one transient nativelike intermediate state I_6 (see Fig. 4 *d*) in the folding process. Folding at larger force (6 pN) is more cooperative than the folding at $f = 0$ pN.

TAR RNA

TAR RNA plays a critical role in HIV viral replication (52). Recent mechanical folding experiments suggested that RNA hairpin folding kinetics can be multistate (20,21,24). Inspired by the biological significance and the intriguing experimental results, we apply our new kinetic model to analyze the folding rates, pathways, and intermediates for TAR RNA here. Unlike the 26-nt hairpin-forming sequences studied above, for the 52-nt TAR RNA sequence, the number of conformations is so large that neither the exact rate matrix/master equation approach nor the kinetic cluster method would be viable. Previous theoretical studies on the folding kinetics for large RNAs mainly use Gō-like model or Monte Carlo simulations (41,51). Our new kinetic model here can give the analytical and stable solution for the populational kinetics folding rates, rate-limiting steps, pathways, and intermediates.

We study the folding kinetics for TAR RNA using the number of parent conformations $\Omega_c = 1$, 10, 20, and 30. The corresponding total number (parent plus daughter) of conformations are 235, 683, 1352, and 1982, respectively. The manageable size of the conformation ensemble allows for the exact solution using the master equation. Fig. 6 shows the folding rate at different forces. We find that the results with different Ω_c values converge for $\Omega_c \geq 10$, and that $\Omega_c = 1$ is not sufficient to give reliable results. This result suggests that having information on the native structure only ($\Omega_c = 1$) might not be sufficient for prediction of the full kinetics.

Folding and unfolding rates. The rate-force (Fig. 6 *a*) curve shows an apparent turn at $f = 12$ pN. For force range ≤ 12 pN, the kinetic process is dominated by the folding reaction. We find that the folding rate decreases as the force increases. For force range ≥ 12 pN, the kinetics is dominated by the unfolding reaction. We find that the unfolding rate increases as the force increases. For force range $f \leq 12$ pN, we can fit our theoretically predicted folding rate versus force $k_r - f$ curve using the analytical formula

$$\ln k_f = 17.9 - 1.8f,$$

which is in good agreement with experimental result (24) $\ln k_f = (21 \pm 8) - (1.92 \pm 0.72)f$. Under a large pulling force $f > f_c = 12$ pN, our predicted unfolding rate k_u leads to the fitted analytical expression

$$\ln k_u = -26.4 + 2.0f,$$

which is again in good agreement with experimental result $\ln k_u = (-24 \pm 16) + (1.68 \pm 1.2)f$.

Folding and unfolding pathways. The populational kinetics (Fig. 6 *b*) solved from the master equation (Eq. S10) based on the (parent plus daughter) conformations shows a kinetic intermediate I_1 at $f = 8$ pN. Fig. S8 shows the secondary structures of the intermediates. To unveil the detailed folding pathway at $f = 8$ pN, we use the kinetic cluster method to classify the 683 conformations (based on the $\Omega_c = 10$ parent states) into kinetic clusters, from which we predict the dominant folding pathways (15,45). Among the 683 conformations, we find six rate-limiting stacks: 2GC3-50GC51, 3CU4-49GG50, 7GG8-45CU46, 17GA18-36UC37, 22GA23-34UC35, and 24GC25-32GC33. The formation of these stacks involves large entropic decrease ΔS, and is thus, slow. Based on the six rate-limiting stacks, we classify the 684 conformations into 35 clusters. Solving the rate matrix for the 35 clusters, we obtain a folding rate of 41 (s^{-1}). From the intercluster transitions (Eq. S12), we identify two major folding pathways from the coil state (I_C) to the native structure (I_N of cluster C_7) (see Fig. 7): $C_1 \rightarrow C_3$ either through C_2 or C'_2 followed by the pathway $\rightarrow C_3 \rightarrow C_4 \rightarrow C_5 \rightarrow C_6 \rightarrow C_7$. Furthermore, Eq. S15 gives a folding rate of 23 s^{-1} for pathway 1 and 17 s^{-1} for pathway 2. The total folding rate is $k_f^{\text{path 1}} + k_f^{\text{path 2}} = 40$ s^{-1}, which is close to 41 s^{-1} from the exact master equation. The rate-limiting steps are $C_1 \rightarrow C_2$ for pathway 1 and $C_1 \rightarrow C'_2$ for pathway 2, corresponding to the slow formation of the native base stacks 22GA23-34UC35 and 24GC25-

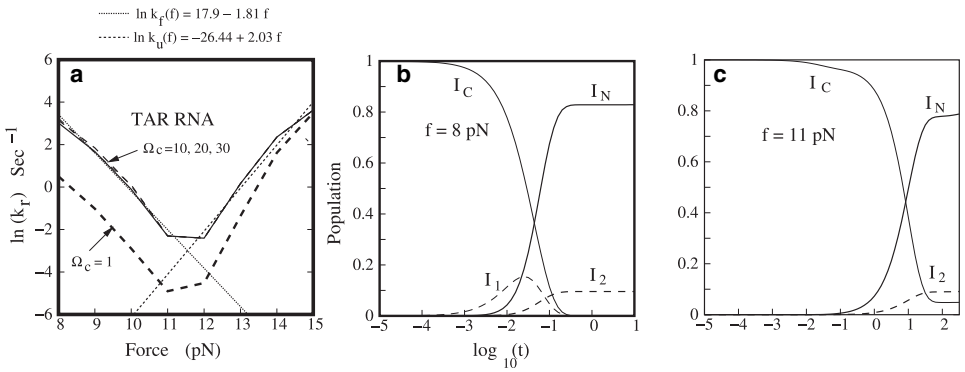

FIGURE 6 (*a*) The calculated folding rates at different forces. The temperature is fixed at 22°C. We use different parameters Ω_c (the number of parent states) to test the convergence of the new kinetic model. We find that the folding rates are identical for $\Omega_c = 10$, 20, and 30. The folding rate from the new kinetic model agrees with the experimental results. We obtain a force-dependence for the folding/unfolding rates as follows: $\ln k_f = 17.9 - 1.8f$ for the folding rate and $\ln k_u = -26.4 + 2.0f$ for the unfolding rate. In the experiment by Li et al. (24), the folding rate is $(21 \pm 8) - (1.92 \pm 0.72)f$ and the unfolding rate is $(-24 \pm 16) + (1.68 \pm 1.2)f$. Therefore, our theoretical prediction is consistent with the experiment. The populational kinetics at different forces: (*b*) $f = 8$ pN and (*c*) $f = 11$ pN.

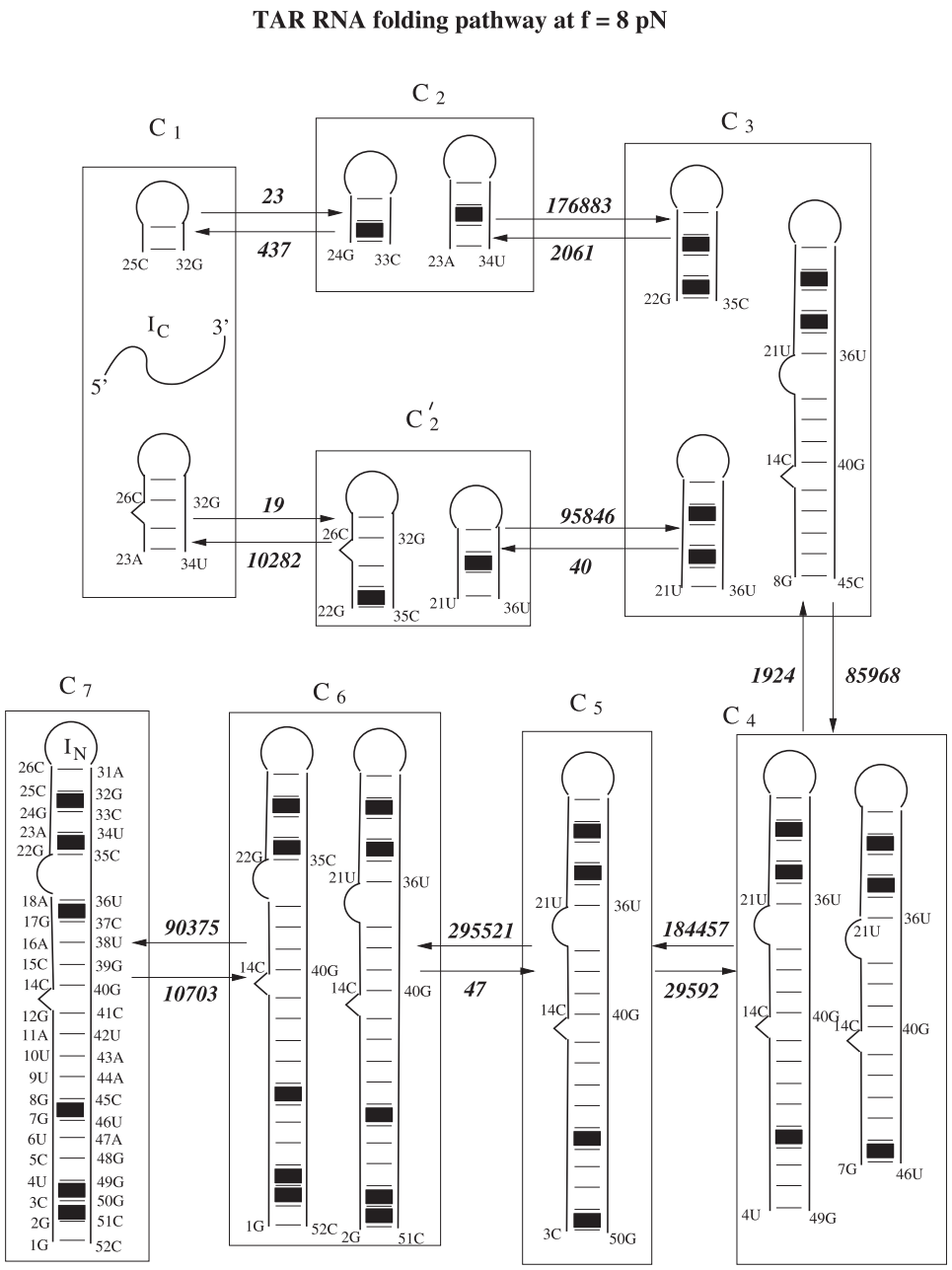

FIGURE 7 The folding of TAR RNA at force $f = 8$ pN and $T = 22°C$ involves two parallel pathways: $C_1 \rightarrow C_2 \rightarrow C_3 \rightarrow C_4 \rightarrow C_5 \rightarrow C_6 \rightarrow C_7$ and $C_1 \rightarrow C'_2 \rightarrow C_3 \rightarrow C_4 \rightarrow C_5 \rightarrow C_6 \rightarrow C_7$. The intercluster transition rate (in s^{-1}) and the dominant micro-pathways between adjacent clusters are shown in the figure. The solid squares denote the rate-limiting base stacks.

32GC33, respectively. The finding is consistent with the experimental results, which suggest that the rate-limiting steps involve the formation of 1.5–5 bp adjacent to the loop depending on the sequence (13).

To further investigate the force effect on the folding kinetics, we compute the populational kinetics at a different force ($f = 11$ pN). We find that the kinetic intermediate I_1 formed at $f = 8$ pN (Fig. 6 *b*) now disappears at a larger force $f = 11$ pN (Fig. 6 *c*). This is due to the decreased stability of structure I_1 at the larger stretching force $f = 11$ pN. Furthermore, we use the kinetic cluster method to investigate the folding pathways. Based on the three slow-forming rate-limiting base stacks: 14CC15-39GG40, 15CA16-38UG39, and 17GA18-36UC37, we classify the original 683 conformations into seven kinetic clusters. The seven-cluster system leads to a folding rate of $0.1\ s^{-1}$, which agrees exactly with the rate constant determined from the master equation for the original complete 683 conformations. The kinetic cluster method gives the dominant folding pathway (Fig. 8) from the coil state (I_C) to the native structure (I_N of cluster C_4). We find that the process is rate-limited by the slow folding from cluster C_1 to cluster C_2. We can estimate the rate for this pathway using Eq. S12 as $p_{I_{(7,1)}} k_{I_{(7,1)} \rightarrow I_{(8,1)}} = (2 \times 10^{-5}\%)(5 \times 10^5 s^{-1}) = 0.1\ s^{-1}$. The slow rate is caused by the very low fractional population (stability) of state $I_{(7,1)}$ in cluster C_1. Because the low stability of $I_{(7,1)}$ comes from the bulge loop between nucleotides 18A and 20C, we conclude

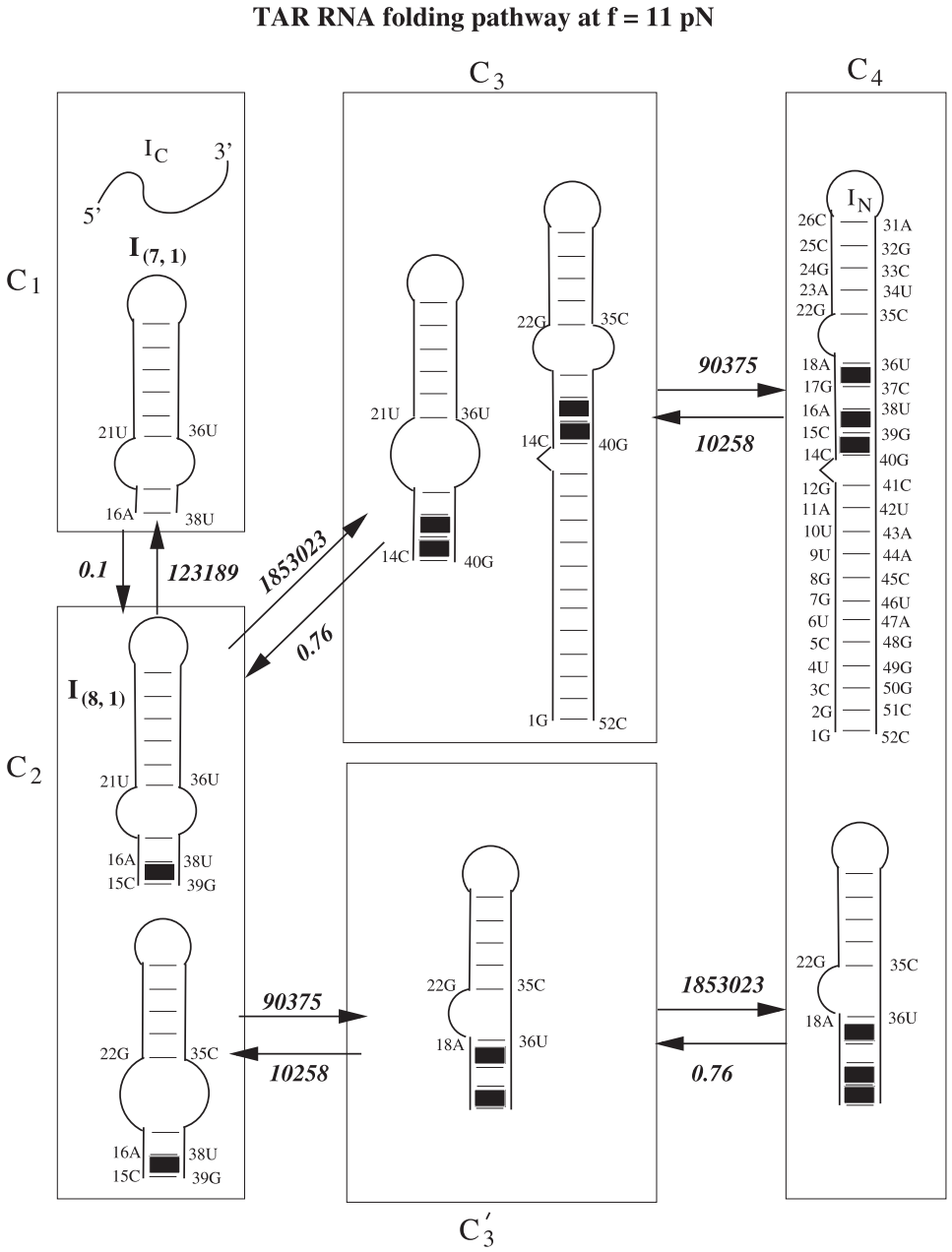

FIGURE 8 The folding of TAR RNA at the transition force (~11 pN) and $T = 22°C$ involves two parallel pathways: $C_1 \rightarrow C_2 \rightarrow C_3 \rightarrow C_4$ and $C_1 \rightarrow C'_2 \rightarrow C_3 \rightarrow C_4$. The intercluster transition rate (in s^{-1}) and the dominant micropathways between adjacent clusters are shown in the figure. The solid squares denote the rate-limiting base stacks.

that the formation of the bulge loop determines the slow folding rate at $f = 11$ pN.

CONCLUSION

Combining the newly developed folding kinetics model with the previously developed kinetic cluster model and the master equation approach, we investigate the effect of the pulling force on RNA folding mechanism such as the folding rate, transition state, kinetic intermediates, and folding pathways for hairpin RNA sequences. This new model, which can predict the misfolded intermediates and pathways, is fundamentally different from the previous Gō-like approaches. The Gō-like model (41) is based only on the native structure and cannot predict the formation of misfolded states in RNA folding. Other simulational methods (36,39,53), which are based on the incomplete sampling of the conformation ensemble, have been useful for the analysis of folding kinetics. However, what distinguishes the new model here from the simulational approaches is the deterministic treatment for the conformational ensemble and the ability to make stable predictions for the long-time folding kinetics.

From our study, we find that the force can significantly alter the kinetic folding pathways, resulting in the change of folding kinetic cooperativity. An increase of the force would turn a noncooperative (non-single-exponential) folding

reaction into a cooperative (single-exponential) kinetics. For the folding process at low force, we find that the rate-limiting step is the formation of only a few (such as two) basepairs adjacent to the loop, which is consistent with the nanomechanical measurements (13). The very low population of the transition state contributes to the slow folding rate. For the unfolding process at high force, we find that the rate-limiting step is to break the base stack of the largest enthalpic barrier. For TAR RNA, using our newly developed kinetic model, we can effectively reduce the number of kinetically accessible conformations to a few hundreds. The folding rate obtained from the new method agrees with the experiment (24) for a wide force regime. In that experiment, the force is >8 pN. We find that the force can change the folding pathways and rate-limiting steps. At $f = 8$ pN, folding involves a single rate-limiting step, namely the formation of the native stack 24GC25-32GC33. At a high force ($f = 11$ pN), the rate-limiting step is the formation of a different native stack, namely, 15CA16-38UG39. The slow folding rate is due to the low stability of the transition state, mainly caused by the formation of a bulge loop. Recent experiments found that the folding process is more complex at low force (20), such as the formation of nonnative base pairs and kinetic trapping in the misfolded state. Further improvement of the model may be required to account for the low-stability misfolded state. In addition, because the approach developed here is independent of any specific structural features, it can be generalized to treat more complex nonhairpin conformations. For more-complex RNA folds, we might need to include a larger number of low-lying parent conformations. Therefore, further calibration of the model may be needed to extend the model to treat general complex RNA folds.

SUPPORTING MATERIAL

Eight figures and 16 equations are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)00597-9.

We thank Dr. Wenbing Zhang for useful discussions. Most of the computations involved in the research for this article were performed on the HPC resources at the University of Missouri Bioinformatics Consortium.

The research was supported by the National Institutes of Health through grant No. GM063732 (to S.-J.C.).

REFERENCES

1. Nagai, K. 1996. RNA-protein complexes. *Curr. Opin. Struct. Biol.* 6:53–61.
2. Vanzi, F., Y. Takagi, H. Shuman, B. S. Cooperman, and Y. E. Goldman. 2005. Mechanical studies of single ribosome/mRNA complexes. *Biophys. J.* 89:1909–1919.
3. Wen, J.-D., L. Lancaster, C. Hodges, A.-C. Zeri, S. H. Yoshimura, et al. 2008. Following translation by single ribosomes one codon at a time. *Nature*. 452:598–603.
4. Ma, H. R., D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua, et al. 2006. Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.* 128:1523–1530.
5. Jung, J. Y., and A. Van Orden. 2006. A three-state mechanism for DNA hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *J. Am. Chem. Soc.* 128:1240–1249.
6. Van Orden, A., and J. Jung. 2008. Fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers.* 89:1–16.
7. Zhang, W. B., and S.-J. Chen. 2002. RNA hairpin folding kinetics. *Proc. Natl. Acad. Sci. USA*. 99:1931–1936.
8. Dill, K. A., S. B. Ozkan, M. S. Shell, and T. R. Weikl. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:289–316.
9. Porschke, D. 1974. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem.* 1:381–386.
10. Ansari, A., S. V. Kunznetsov, and Y. Shen. 2001. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl. Acad. Sci. USA*. 98:7771–7776.
11. Wallace, M. I., L. Ying, S. Balasubramanian, and D. Klenerman. 2001. Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc. Natl. Acad. Sci. USA*. 98:5584–5589.
12. Bonnet, G., O. Krichevsky, and A. Libchaber. 1998. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci. USA*. 95:8602–8606.
13. Woodside, M. T., W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, et al. 2006. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. USA*. 103:6190–6195.
14. Woodside, M. T., G. García-García, and S. M. Block. 2008. Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.* 12:640–646.
15. Zhang, W. B., and S.-J. Chen. 2006. Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophys. J.* 90:765–777.
16. Tinoco, Jr., I. 2004. Forces as a useful variable in reactions: unfolding RNA. *Annu. Rev. Biophys. Biomol. Struct.* 33:363–385.
17. Tinoco, Jr., I., P. T. X. Li, and C. Bustamante. 2006. Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.* 39:325–360.
18. Zhuang, X., and M. Rief. 2003. Single-molecule folding. *Curr. Opin. Struct. Biol.* 13:88–97.
19. Bustamante, C., S. B. Smith, J. Liphardt, and D. Smith. 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* 10:279–285.
20. Li, P. T. X., C. Bustamante, and I. Tinoco, Jr. 2007. Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc. Natl. Acad. Sci. USA*. 104:7039–7044.
21. Li, P. T. X., J. Vieregg, and I. Tinoco, Jr. 2008. How RNA unfolds and refolds. *Annu. Rev. Biochem.* 77:77–100.
22. Chen, G., J.-D. Wen, and I. Tinoco, Jr. 2007. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA*. 13:2175–2188.
23. Green, L., C.-H. Kim, C. Bustamante, and I. Tinoco, Jr. 2008. Characterization of the mechanical unfolding of RNA pseudoknots. *J. Mol. Biol.* 375:511–528.
24. Li, P. T. X., D. Collin, S. B. Smith, C. Bustamante, and I. Tinoco, Jr. 2006. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* 90:250–260.
25. Wen, J. D., M. Manosas, P. T. X. Li, S. B. Smith, C. Bustamante, et al. 2007. Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys. J.* 92: 2996–3009.
26. Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, Jr., and C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. *Science*. 292:733–737.
27. Liphardt, J., S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante. 2002. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*. 296:1832–1835.

28. Onoa, B., S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco, Jr., et al. 2003. Identifying kinetic barriers to mechanical unfolding of the *T. thermophilia* ribozyme. *Science*. 299:1892–1895.

29. Harlepp, S., T. J. Marchal, J. Robert, F. Leger, A. Xayaphoumimine, et al. 2003. Probing complex RNA structures by mechanical force. *Eur. Phys. J. E.* 12:605–615.

30. Gerland, U., R. Bundschuh, and T. Hwa. 2001. Force induced denaturation of RNA. *Biophys. J.* 81:1324–1332.

31. Gerland, U., R. Bundschuh, and T. Hwa. 2003. Mechanically probing the folding pathway of single RNA molecules. *Biophys. J.* 84: 2831–2840.

32. Cocco, S., R. Monasson, and J. F. Marko. 2001. Force and kinetic barriers of unzipping of the DNA double helix. *Proc. Natl. Acad. Sci. USA*. 98:8608–8613.

33. Liu, F., H. Tong, and Z. C. Ou-yang. 2006. Force unfolding single RNAs. *Biophys. J.* 90:1895–1902.

34. Manosas, M., J.-D. Wen, P. T. X. Li, S. B. Smith, C. Bustamante, et al. 2007. Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. *Biophys. J.* 92:3010–3021.

35. Ritort, F. 2006. Single-molecule experiments in biological physics: methods and applications. *J. Phys. Condens. Matter*. 18:R531–R583.

36. Hyeon, C., and D. Thirumalai. 2005. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. USA*. 102:6789–6794.

37. Hyeon, C., and D. Thirumalai. 2007. Mechanical unfolding of RNA: from hairpins to structures with internal multiloops. *Biophys. J.* 92: 731–743.

38. Hyeon, C., G. Morrison, and D. Thirumalai. 2008. Force-dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes. *Proc. Natl. Acad. Sci. USA*. 105: 9604–9609.

39. Yeh, I.-C., and G. Hummer. 2004. Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations. *Biophys. J.* 86:681–689.

40. Bustamante, C., J. F. Marko, E. D. Siggia, and S. B. Smith. 1994. Entropic elasticity of λ-phage DNA. *Science*. 265:1599–1600.

41. Cocco, S., J. F. Marko, and R. Monasson. 2003. Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur. Phys. J. E.* 10:153–161.

42. Cao, S., and S.-J. Chen. 2008. Predicting ribosomal frameshifting efficiency. *Phys. Biol.* 5:016002.

43. Chen, S.-J. 2008. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* 37:197–214.

44. Konishi, Y., T. Ooi, and H. A. Scheraga. 1982. Regeneration of ribonuclease A from the reduced protein. Rate-limiting steps. *Biochemistry*. 21:4734–4740.

45. Cao, S., and S.-J. Chen. 2007. Biphasic RNA folding kinetics and telomerase activity. *J. Mol. Biol.* 367:909–927.

46. Geis, M., C. Flamm, M. T. Wolfinger, A. Tanzer, I. L. Hofacker, et al. 2008. Folding kinetics of large RNAs. *J. Mol. Biol.* 379:160–173.

47. Flamm, C., and I. L. Hofacker. 2008. Beyond energy minimization: approaches to the kinetic folding of RNA. *Monatsh. Chem.* 139:447–457.

48. Tang, X., S. Thomas, L. Tapia, D. P. Giedroc, and N. M. Amato. 2008. Simulating RNA folding kinetics on approximated energy landscapes. *J. Mol. Biol.* 381:1055–1067.

49. Serra, M. J., and D. H. Turner. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol.* 259:242–261.

50. Rief, M., H. Clausen-Schaumann, and H. E. Gaub. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 6:346–349.

51. Hyeon, C., and D. Thirumalai. 2008. Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins. *J. Am. Chem. Soc.* 130:1538–1539.

52. Vrolijk, M. M., M. Ooms, A. Harwig, A. T. Das, and B. Berkhout. 2008. Destabilization of the TAR hairpin affects the structure and function of the HIV-1 leader RNA. *Nucleic Acids Res.* 36:4352–4363.

53. Koplin, J., Y. Mu, C. Richter, H. Schwalbe, and G. Stock. 2006. Structure and dynamics of an RNA tetraloop: a joint molecular dynamics and NMR study. *Structure*. 13:1255–1267.