

Conformational entropy of a pseudoknot polymer

Yu-Jane Sheng^{a)} and You-Chin Mou

Department of Chemical Engineering, National Taiwan University, Taipei, Taiwan 106, Republic of China

Heng-Kwong Tsao^{b)}

Department of Chemical and Materials Engineering, National Central University, Zhongli, Taiwan 320, Republic of China

(Received 7 November 2005; accepted 6 January 2006; published online 27 March 2006)

The thermodynamics and kinetics of *ABAB* pseudoknot formation owing to reversible intrachain reactions are investigated for a flexible polymer based on the off-lattice Monte Carlo simulations. The polymer is made of N hard spheres tethered by inextensible bonds and consists of two reactive pairs *AA* and *BB* with binding energies $-\varepsilon_1$ and $-\varepsilon_2$, respectively, and three loop lengths (ℓ_1 , ℓ_2 , and ℓ_3). Although two intermediate states, loops *A* and *B*, may be formed, the folding path goes mainly through the intermediate loop whose free energy reduction associated with coil-to-loop crossover is greater. The conformational entropy loss is found to follow $\Delta S = \alpha \ln N + G$, where $\alpha \approx 2.48$ for coil-loop crossover and $\alpha \approx 2.43$ for loop-pseudoknot crossover. The constant G depends on the three loop lengths and the two end-to-reactive site lengths (L_1 and L_2). For a given total loop length, G is maximum when the three loop lengths are equal ($\ell_1 = \ell_2 = \ell_3$). When $\ell_1 = \ell_3$, the entropy loss is minimum if $\ell_2 = 0$. However, the condition $\ell_1 \neq \ell_3$ makes G even smaller. This consequence indicates that asymmetry in loop lengths is thermodynamically favorable and this fact is consistent with observations of pseudoknotted RNA structures. © 2006 American Institute of Physics.

[DOI: [10.1063/1.2172600](https://doi.org/10.1063/1.2172600)]

I. INTRODUCTION

Biopolymers such as RNA can fold into a topological structure called a pseudoknot. As shown in Fig. 1, a pseudoknot is a loop in which one end of the chain returns and forms additional contact with the loop. Thus, a pseudoknot polymer contains two intrachain contacts (stems) and two loops. Unlike true knots, pseudoknots are the tertiary structures of the folded RNA's and are shown to exist in almost all RNA classes including transfer, messenger, ribosomal, viral, catalytic, and self-splicing RNA's.¹ In addition to stabilizing the native fold, pseudoknot is believed to be essential for various functions, such as mediating the binding of the proteins they encode² and labeling functionally important positions on the coding region of the mRNA sequence.³

Pseudoknots are not true knots in the conventional sense.⁴⁻⁶ It is simply generated by a linear polymer's self-contacts and thereby is encountered more frequently. From a computational viewpoint, a pseudoknot is interesting owing to the fact that it can be regarded as the simplest tertiary structure.⁷ For a sequence of length N , the dynamic programming algorithm for predicting optimal RNA secondary structure requires $O(N^4)$ in time for simple loop conformations. Including pseudoknots increases the conformational complexity to $O(N^6)$.⁸ The folded molecule such as proteins is even more complicated, involving an exponential scaling, a^N .

Predicting the secondary structure of a RNA or single-stranded DNA molecules with minimum free energy has

been extensively investigated for the past two decades.⁹⁻¹³ In the widely used nearest-neighbor thermodynamic model, the free energies are assigned to loops rather than to base pairs.^{9,10} The base-pairing graph for a molecule is decomposed into distinct loops that are associated with empirically measured enthalpic and entropic terms. Canonical loops include stacked based pairs, hairpin loop, bugle loop, internal loop, etc. Thus, the total free energy is the sum of free energies of all loops in the secondary structure. The free energy parameters depend on loop sequence, length, and type.

For classical RNA secondary structure, relatively few experiments are available for loop structures and thereby extrapolations based on the polymer theory for loop thermodynamics are adopted to give the free energy parameters. For example, the length dependence of the free energy increment is $\Delta G = \alpha k_B T \ln N$, where N is the number of unpaired nucleotides in the loop and $\alpha \approx 1.75$ according to estimation based on self-avoiding lattice chains.¹⁴ While thermodynamic parameters for nonpseudoknotted secondary structure are evaluated with satisfactory accuracy,¹¹ systematic studies on pseudoknot thermodynamics are not available. Recently, attempts to estimate the free energy parameters for the simplest H-pseudoknots, also based on the general theory of polymer loop thermodynamics, are presented.^{11,12}

The approaches based on a search for a state of global free energy minimum are equilibrium methods and do not consider the folding pathways. However, some RNA's, particularly long molecules, do not adopt the structure with the lowest free energy. The differences between real structures and the minimum energy states may be caused by the existence of specific folding pathways capturing the molecules in

^{a)}Electronic mail: yjsheng@ntu.edu.tw

^{b)}Electronic mail: hktsao@cc.ncu.edu.tw

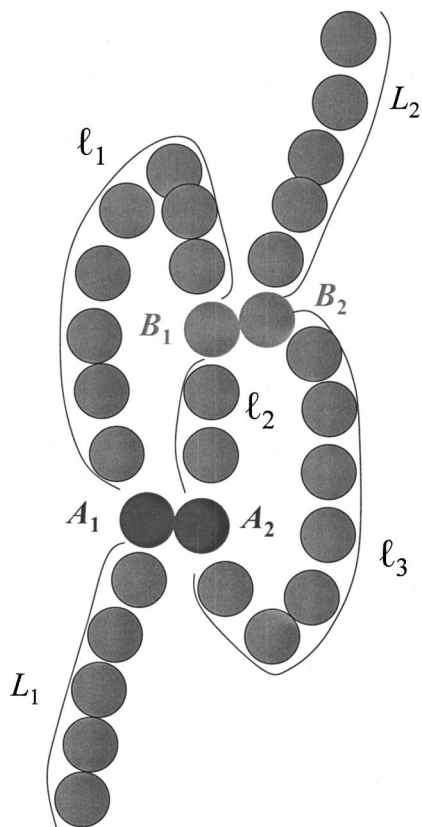


FIG. 1. The schematic of a pseudoknot polymer. The two pairs of reactive groups located at beads $\{A_1, A_2\}$ and $\{B_1, B_2\}$.

local minima.¹⁰ Consequently, simulating RNA-folding dynamics can provide an alternative avenue to predict pseudoknot in the long time limit, in addition to bring some unique insights into the kinetic aspects of RNA folding.¹⁵ However, a recent determination of the simple two-state kinetics associated with loop-to-coil crossover of a ssDNA molecular beacon indicates the formidable challenges of measuring these rates in real systems.¹⁶ One cannot anticipate rapid progress in modeling folding kinetics when it remains so difficult to determine the underlying parameters.¹⁰ Studying kinetics and thermodynamics of pseudoknot formation within the homopolymer context may provide valuable insights into understanding of physics of RNA pseudoknot. The pioneering work in the prediction of the folding thermodynamics of RNA with pseudoknot formation was primarily based on the homopolymer transversing a random walk on the lattice.^{5,7,17,18} The thermal transition of pseudoknot polymers to their denatured states was reported through a first-order transition,¹⁷ a second-order transition,¹⁸ or continuous transition.⁷

The free energy of a pseudoknot structure is primarily the sum of the free energies of stacking in the stem (stabilizing negative value) and thermal fluctuation in the loop (destabilizing positive value).¹¹ The stacking energy can be evaluated using the known nearest-neighbor model parameter of helix propagation. For the loop energies, however, crude models based on the polymer theory were employed.^{11,12} Assuming that loops have pure entropic nature, the free energy of formation of a loop of N nucleotides

is modeled by Gaussian chains. The main limitation of this structural model is the absence of excluded volume effect between nearby single strands and stems.¹² In the present study, we investigate the folding kinetics and thermodynamics of a pseudoknotted polymer by off-lattice Monte Carlo simulations. In order to avoid complications caused by attractions among monomers and solvents, we aim to study a homopolymer with only hard-core excluded volume effects. Since we will focus mainly on the conformational entropy associated with pseudoknot, the stem i is simply modeled as a pair of reactive beads with binding energy $-\epsilon_i$. The conformational entropy loss associated with pseudoknot formation is quantified for a flexible polymer with four reactive sites $ABAB$. The variation of the entropy loss with the distribution of loop lengths can then be compared with the observation of pseudoknotted RNA structures by the Pseudoknot database.¹⁹

II. THEORY

A. Thermodynamic approach

Consider a flexible chain of N hard spheres with total length L . The reactive groups are located at beads $A_1, B_1, A_2,$ and B_2 . The four reactive sites divide the chain into five blocks with lengths $\{L_1, \ell_1, \ell_2, \ell_3, L_2\}$, as illustrated in Fig. 1. The reactive sites A_1 and B_2 are positioned at distance L_1 and L_2 , respectively, from the chain ends. The contour distances between the active groups are $\ell_1(A_1-B_1)$, $\ell_2(B_1-A_2)$, and $\ell_3(A_2-B_2)$. When the two reactive sites A_1 and A_2 (B_1 and B_2) are in close proximity (contact), the reversible reaction occurs and the binding energy is $-\epsilon_1(-\epsilon_2)$. The partition function associated with such a pseudoknot polymer depends on the location of the reactive sites and is given by

$$\mathcal{Z}\left(\frac{L_1}{L}, \frac{\ell_1}{L}, \frac{\ell_2}{L}, \frac{\ell_3}{L}; N\right) = \sum_i g_i \exp(\beta \epsilon_i), \quad (1)$$

where g_i denotes the degeneracy associated with the state i of internal energy $-\epsilon_i$ and β is the inverse temperature, $\beta = (k_B T)^{-1}$.

When only one binding pair AA (BB) is formed, the conformation is identified as the loop state l_A (l_B). Formation of two binding pairs corresponds to the pseudoknot state (p). Otherwise, the polymer conformations are regarded as the coil state (c). As a result, the partition function can be divided into four parts,

$$\mathcal{Z} = \mathcal{Z}_c + \mathcal{Z}_{l_A} + \mathcal{Z}_{l_B} + \mathcal{Z}_p, \quad (2)$$

where $\mathcal{Z}_c = g_c$, $\mathcal{Z}_{l,k} = g_{l,k} \exp(\beta \epsilon_k)$, and $\mathcal{Z}_p = g_p \exp[\beta(\epsilon_1 + \epsilon_2)]$. Evidently, the degeneracy g_i varies with N and the distribution of block lengths. In terms of the partition function, the probability of the state i is simply

$$P_i(T) = \frac{\mathcal{Z}_i}{\mathcal{Z}} = \frac{K_i e^{\beta \epsilon_i}}{1 + \sum_i K_i e^{\beta \epsilon_i}}, \quad (3)$$

where $K_i = g_i / g_c$ and $i = c, A(\text{loop}), B(\text{loop}),$ and p . The probability curve $P_j(T)$ represents the variation of the state j with temperature. Since the conformational entropy at the state j is given by $S_j = k_B \ln g_j$, K_j is related to the conformational entropy loss $\Delta S_{c,j}$ from coil to j state by

$$\Delta S_{c,j} = S_j - S_c = k_B \ln K_j. \quad (4)$$

Equation (4) presents an important consequence that one is able to evaluate the entropy change of a polymer chain from coil-to-loop or pseudoknot state from the equilibrium constant. If the probability curves, $P_i(T)$, are obtained from Monte Carlo simulations, the constant K_i may be extracted by curve fitting.²⁰

B. Kinetic approach

Since the binding energy ε_i is comparable to the thermal energy $k_B T$ in the present study, the formation of loop or pseudoknot is reversible. As a consequence, the coil-to-pseudoknot crossover can be regarded as chemical equilibrium associated with four reversible reactions. The reactant, intermediates, and product are, respectively, the coil, loops A and B , and pseudoknot states. There are four species in two possible pathways:

path I: coil (c) \leftrightarrow loop $AA(l_A)$ \leftrightarrow pseudoknot (p),

path II: coil (c) \leftrightarrow loop $BB(l_B)$ \leftrightarrow pseudoknot (p).

Note that the reaction can be regarded as an irreversible one if $\beta\varepsilon_i \gg 1$. At chemical equilibrium, the principle of detailed balance must be satisfied. The characteristics of the kinetics can then be studied by examining the rate constants evaluated from Monte Carlo simulations.²⁰⁻²²

Following the Arrhenius kinetics, the rate coefficients $k_{i,j} = \tau_{i,j}^{-1}$ from the i to j species is given by

$$k_{i,j} = k_{i,j}^0 \exp(-\beta F_{i,j}), \quad (5)$$

where $F_{i,j}$ denotes the free energy barrier associated with changing from the i to j species. The preexponential factor $k_{i,j}^0$ is temperature independent. When the conformation of a hard-sphere chain varies from the coil-to-loop state or from the loop-to-pseudoknot state, it is anticipated that the free energy barrier originates from the entropy loss. That is,

$$\beta F_{c,l} = -\Delta S_{c,l}/k_B \text{ and } \beta F_{l,p} = -\Delta S_{l,p}/k_B. \quad (6)$$

On the other hand, as the polymer conformation fluctuates from the loop-to-coil state or from the pseudoknot-to-loop state, the free energy barrier, which has to be overcome, is simply the binding energy, $\beta F_{l,c} = \beta\varepsilon_1$ or $\beta F_{p,l} = \beta\varepsilon_2$. When chemical equilibrium is reached, the detailed balance, $k_{i,j}P_i = k_{j,i}P_j$, also yields Eq. (4) with the equilibrium constant K_i defined as

$$K_1 = \frac{k_{c,l_A}^0}{k_{l_A,c}^0} \exp\left(\frac{\Delta S_{c,l_A}}{k_B}\right), \quad K_2 = \frac{k_{c,l_B}^0}{k_{l_B,c}^0} \exp\left(\frac{\Delta S_{c,l_B}}{k_B}\right), \quad (7)$$

$$K_3 = \frac{k_{l_B,p}^0}{k_{p,l_B}^0} \exp\left(\frac{\Delta S_{l_B,p}}{k_B}\right), \quad K_4 = \frac{k_{l_A,p}^0}{k_{p,l_A}^0} \exp\left(\frac{\Delta S_{l_A,p}}{k_B}\right).$$

The principle of detailed balance requires $K_1 K_4 = K_2 K_3$ and $k_{i,j}^0 = k_{j,i}^0$. If one calculates the inverse of the rate constant from Monte Carlo simulations, then the equilibrium constants K_i or entropy loss can be obtained directly.

III. MONTE CARLO SIMULATION

In this paper, off-lattice Monte Carlo (MC) simulations were performed to study both thermodynamics and kinetics associated with coil-to-pseudoknot crossover. In order to calculate the conformational entropy change associated with the formation of pseudoknot polymers by the aforementioned theory, we study polymers with only hard-core excluded volume effects but without other extra interaction energy such as attractions among monomers and solvent. The model chain is made of N hard spheres of diameter σ with two pairs of attractive beads $A_1 A_2$ and $B_1 B_2$. This polymer is a freely jointed chain in continuous space. The interactions between the bonded beads are through the infinite deep square-well potentials,²⁰⁻²²

$$U_{i,i+1} = \begin{cases} \infty, & r < \sigma \\ 0, & \sigma \leq r < \xi\sigma \\ \infty, & r \geq \xi\sigma, \end{cases}$$

where $\xi=1.2$. Bond crossing (phantom chain) can be prevented by such a choice. For this model associating chain, the interaction between attractive beads A_1 and A_2 is represented by a standard square-well potential,

$$U_{A,A} = \begin{cases} \infty, & r < \sigma \\ -\varepsilon_1, & \sigma \leq r < \lambda\sigma \\ 0, & r \geq \lambda\sigma, \end{cases}$$

where $\lambda=1.2$. Similarly, we adopt $-\varepsilon_2$ for $U_{B,B}$, which depicts attraction between B_1 and B_2 . Without the loss of generality, we assume the binding energy $\varepsilon_1=10\varepsilon$ and $\varepsilon_2=25\varepsilon$.

Monte Carlo simulation is generally not the method of choice for investigating kinetics. When the move of the particle is purely local, however, the dynamical interpretation of MC is reasonable.^{22,23} Moreover, since the actual dynamics has reached thermodynamic equilibrium for those reversible reactions, it is justified to estimate the reaction constants of the kinetics by MC method. The chain is identified as in the ‘‘coil’’ state when the attractive beads of the chain are not within square-well interaction regimes. That is, $|\mathbf{r}_i - \mathbf{r}_j| > \lambda\sigma$. The ‘‘loop’’ state is defined as the formation of only one attractive pair such as $A_1 A_2$ or $B_1 B_2$. That is, $|\mathbf{r}_{A_1} - \mathbf{r}_{A_2}| < \lambda\sigma$. When both attractive pairs are formed, the polymer is regarded as being in ‘‘pseudoknot’’ state. The rate constant $k_{i,j}$ is evaluated from the inverse of the mean time $\tau_{i,j}$. It is defined as the mean period of staying at the state i , which jumps to the state j later. The probability of the state i is calculated by the total Monte Carlo steps (MCSs) staying at the state i to the total MCSs of simulation. The detailed balance, $k_{i,k}P_i = k_{k,i}P_k$, is fulfilled in our simulations.

The systems simulated contain an isolated polymer chain with chain length N ranging from 19 to 59. The simulations are performed under the conditions of constant temperature and total number of beads. The reduced temperature $T^* = T/\varepsilon$ is varied to obtain the probability curves for coil, loop, and pseudoknot states. The trial moves employed for chains of the equilibration and production process are bead displacement motions. They involve randomly picking a bead and displacing it to a new position in the vicinity of the

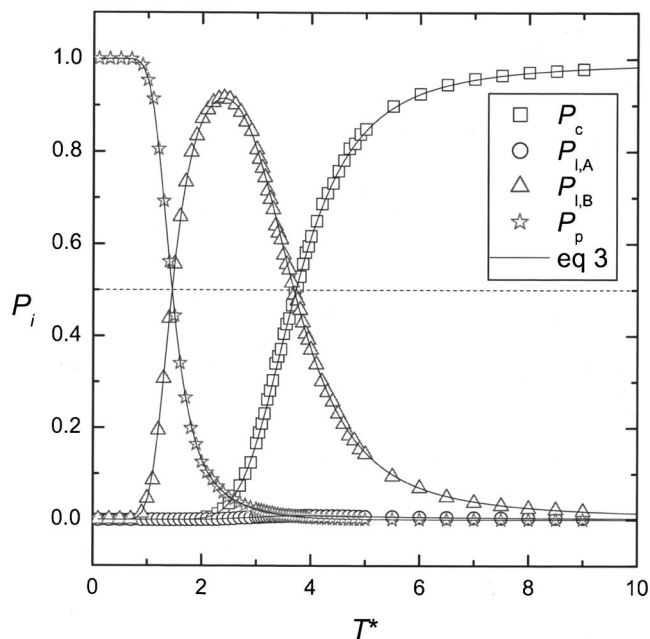


FIG. 2. Typical probability curves for chain length $N=19$. The MC results and theoretical curves, Eq. (3), are denoted by data points and solid curves, respectively.

old position. The distance away from the original position is chosen with a probability, which satisfies the condition of equal sampling of all points in the spherical shell surrounding the initial position. The new configurations resulting from this move are accepted according to the standard Metropolis acceptance criterion. Runs for the same chain length at different temperatures are performed starting with the final configuration from a previous temperature and are equilibrated for 2×10^8 steps. Measurements for static properties such as the probabilities of coil or pseudoknot states are taken over a period of $(5-10) \times 10^6$ MCSs per bead. Note that the chain relaxation time defined by end-to-end autocorrelation function ($\varepsilon_i=0$) is about 10^5 MCSs per bead for $N=20$.

IV. RESULTS AND DISCUSSION

The folding process from a random coil to a pseudoknot is investigated. On the basis of the four-state model (coil, loop A, loop B, and pseudoknot), both the thermodynamic and kinetic approaches are employed to yield consistent results. By Monte Carlo simulations, the entropy loss associated with the coil-to-pseudoknot crossover $\Delta S_{c,p}$ is calculated through the equilibrium constant K_i in the probability curves, Eq. (3). Because of four reactive sites, there are five blocks characterized by lengths $\{L_i, \ell_i\}$. Evidently, the number of coil conformations varies only with the chain length, $g_c(N)$, while the numbers of loop and pseudoknot conformations also depend on the locations of reactive sites L_i/L , ℓ_i/L , i.e., $g_p(L_i/L, \ell_i/L, N)$. Although there are many combinations of the block lengths, we focus on two cases in this study. In the first case the effect of chain length on the pseudoknot entropy is studied for equal block length, i.e.,

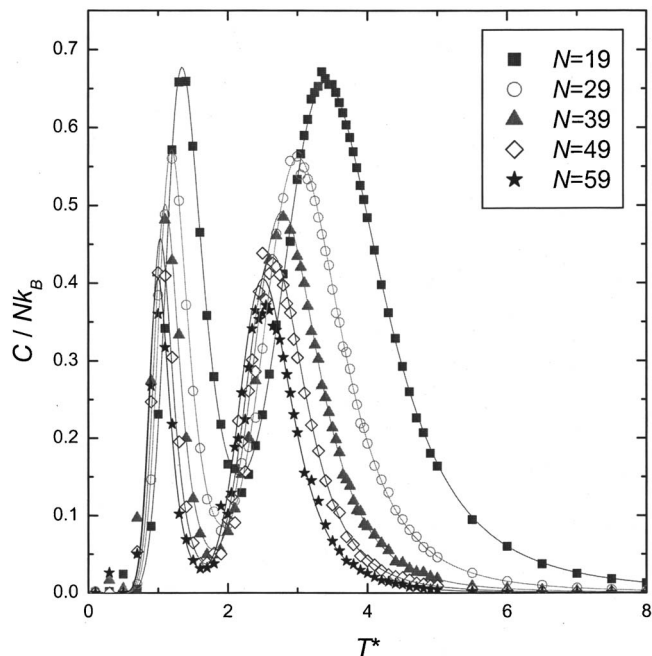


FIG. 3. The heat capacity curves for different chain lengths are plotted with C/Nk_B as a function of T^* and compared to Eq. (9).

$L_i = \ell_i$. In the second case, how the distribution of the loop lengths (ℓ_i) influences the pseudoknot entropy is examined by fixing the chain length N and end block lengths $L_1=L_2$.

A. Crossover temperatures and folding path

The probability curves for various states were evaluated by performing off-lattice Monte Carlo simulations. The simulation results can be well described by the coil-loop-pseudoknot model, Eq. (3). Figure 2 depicts a typical example of P_i for a given chain length $N=19$ with $L_i=\ell_i=3$. All probability curves are excellently represented by Eq. (3) with the equilibrium constants K_i estimated from the rate constants $k_{i,j}$, which are directly obtained from MC simulations. As anticipated, the coil state prevails at higher temperature ($T^* > T_c^*$) while the pseudoknot state dominates at lower temperature ($T^* < T_p^*$). Here T_c^* is defined as the temperature at which $P_c=1/2$ and similarly, T_p^* is for

TABLE I. The variation of the characteristic temperatures T^* with the chain length N based on (a) probability curves $P_i(T)$ and (b) heat capacity curves $C(T)$.

(a)	T_p^*	T_l^*	T_c^*
$N=19$	1.45	2.42	3.72
$N=29$	1.27	2.11	3.20
$N=39$	1.16	1.92	2.94
$N=49$	1.09	1.83	2.77
$N=59$	1.05	1.75	2.65
(b)	T_p	T_l	T_c
$N=19$	1.34	2.16	3.42
$N=29$	1.20	1.92	3.02
$N=39$	1.10	1.78	2.78
$N=49$	1.04	1.70	2.64
$N=59$	1.00	1.62	2.50

$P_p = 1/2$. In the intermediate temperature ($T_p^* < T^* < T_c^*$), however, one observes that the loop B state is dominant. In other words, the folding process goes primarily through path II instead of path I.

The coil-to-pseudoknot crossover involves the formation of the intermediate state, loop A or B state. The probability ratio of path II to path I can be obtained from Eq. (3),

$$\frac{\text{path II}}{\text{path I}} = \frac{P_B}{P_A} = \frac{K_2 e^{\beta \varepsilon_2}}{K_1 e^{\beta \varepsilon_1}}. \quad (8)$$

According to Eq. (4), the equilibrium constant K_i is related to the entropy change $\Delta S_{c,i}$. Therefore, the probability ratio is determined by the overall free energy change associated with the coil-to-loop crossover ΔG ,

$$\frac{\text{path II}}{\text{path I}} = \exp[-\beta(\Delta G_B - \Delta G_A)], \quad (9)$$

where $\Delta G_i = (-\varepsilon_i) - T\Delta S_{c,i}$. The dominant pathway is the one with more total free energy reduction. When $\Delta S_{c,iA} \approx \Delta S_{c,iB}$ and $\exp[\beta(\varepsilon_2 - \varepsilon_1)] \gg 1$, one has $P_B \gg P_A$ and path II dominates.

The characteristics of coil-to-pseudoknot crossover can also be illustrated by the heat capacity curve, which depicts the variation of the heat capacity with temperature, $C(T)$. Since the heat capacity is related to the internal energy U by

$$C = \frac{\langle U^2 \rangle - \langle U \rangle^2}{k_B T^2},$$

the heat capacity curve can be evaluated either from MC simulations or directly from Eq. (3),

$$\frac{C}{Nk_B} = \frac{[P_A \varepsilon_1^2 + P_B \varepsilon_2^2 + P_p(\varepsilon_1 + \varepsilon_2)^2] - [P_A \varepsilon_1 + P_B \varepsilon_2 + P_p(\varepsilon_1 + \varepsilon_2)]^2}{N(k_B T)^2}. \quad (10)$$

Figure 3 shows the heat capacity curves for various chain lengths. There are three extremes, $\partial C/\partial T = 0$, corresponding to three characteristic temperatures $\{T_c, T_l, T_p\}$. As shown in Table I, they can be approximately represented by $\{T_c^*, T_l^*, T_p^*\}$ from probability curves, where $P_c(T_c^*) = P_p(T_p^*) = 1/2$ and T_l^* corresponding to the maximum probability of the loop B state, $(\partial P_l/\partial T)_{T_l^*} = 0$. Evidently, $(\partial^2 C/\partial T^2)_{T_c} < 0$ and $(\partial^2 C/\partial T^2)_{T_p} < 0$ denote the coil-to-loop and loop-to-pseudoknot crossovers, respectively. On the other hand, $(\partial^2 C/\partial T^2)_{T_l} > 0$ indicates the stability of the intermediate loop state. As shown in Table I, T_c , T_p , and T_l decline with increasing the chain length N . The characteristic temperature associated with the i to j crossover can be estimated by letting $\Delta G_{i,j} = 0$, where the internal energy gain is balanced by the entropy loss,

$$T_{i,j}^* \approx \frac{\Delta U_{i,j}}{\Delta S_{i,j}}.$$

For the same internal energy gain $-\varepsilon_i$, the conformation entropy loss rises with increasing the chain length. As a result, the crossover temperature decreases with N . If the heat capacity curve is approximately symmetry in the range $T_p < T < T_c$, the characteristic temperature T_l can be estimated by the arithmetic mean of the crossover temperatures, $T_l \approx (T_c + T_p)/2$.

B. Rate constants and conformational entropy

In the present study, the loop or pseudoknot formation is reversible and the rate constants are different from those associated with irreversible loop²⁴ or pseudoknot formation. The latter case corresponds to that a random coil ($T \gg T_c$) is suddenly quenched in an environment favoring pseudoknot

formation ($T \ll T_p$). The free energy barrier of the first passage time is simply the entropy loss, which is much less than the binding energy gain, i.e., $\beta \varepsilon_i \rightarrow \infty$.

The unfolding kinetics is dominated by the binding energy. That is, the thermal fluctuations provide a probability of $\exp(-\beta \varepsilon_i)$ to unbind the pseudoknot or loop conformation regardless of the chain length. When $\ln k_{p,l}^{-1}$ or $\ln k_{l,c}^{-1}$ is plotted against the inverse temperature β , all the rate coefficients evaluated from different chain lengths for loop B -to-coil and pseudoknot-to-loop A crossovers collapse into a single line with a slope of the binding energy $\beta \varepsilon_2$. Similarly, another line with the slope $\beta \varepsilon_1$ is obtained for pseudoknot-to-loop B and loop A -to-coil crossovers. Figure 4(a) confirms that the rate constants associated with unfolding kinetics are independent of chain length N . The scattering data of $k_{p,A}$ is due to the small probability associated with path I.

On the other hand, the folding kinetics is controlled by the conformational entropy. Since $\beta F_{i,j} = -\Delta S_{i,j}/k_B$, the rate constants are independent of temperature but vary with chain lengths. As illustrated in Fig. 4(b) for coil-to-loop B and Fig. 4(c) for loop B -to-pseudoknot folding kinetics, horizontal lines are obtained in the plot of $\ln k_{c,l}^{-1}$ and $\ln k_{l,p}^{-1}$ against the inverse temperature β for various chain lengths. Although loop B formation is preferred than loop A formation, the entropy loss from coil to either loop A or loop B should be equivalent owing to the same loop conformation for a polymer with equal block lengths. Therefore, the significantly different probabilities between them ($P_{l,A}$ and $P_{l,B}$) are caused by the open (loop \rightarrow coil) kinetics. The larger binding energy renders the loop B state thermodynamically favorable.

Equation (5) indicates that $\ln k_{i,j}^{-1}$ for folding kinetics is proportional to the change of conformational entropy, $-\Delta S_{i,j}$. Accordingly, the result that the intercept is increased with

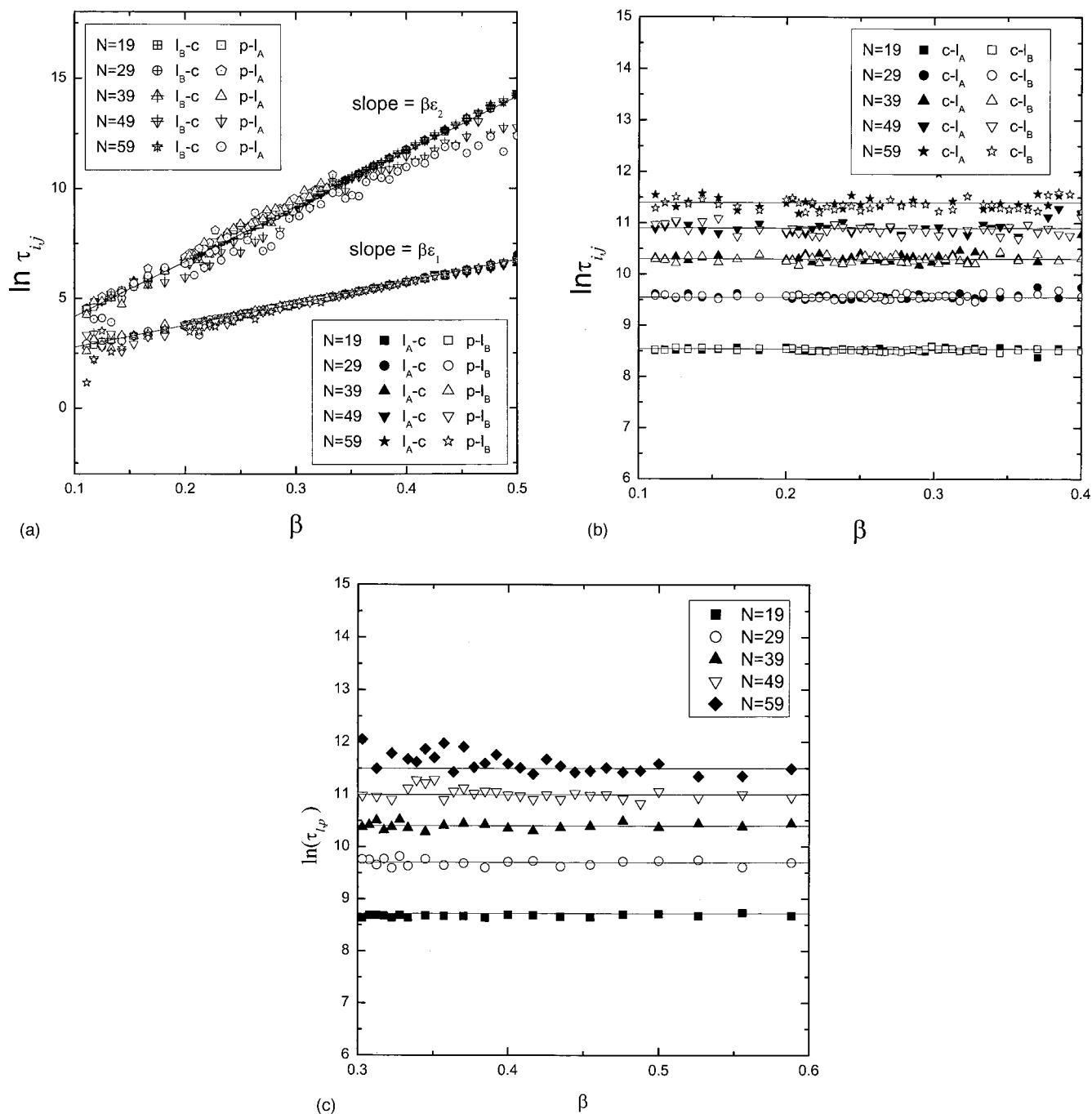


FIG. 4. (a) The variation of the rate constants $k_{l,c}$ and $k_{p,l}$ with the inverse temperature β for different chain lengths. The slopes of the solid line are $\beta \epsilon_1 = 10$ for $p \rightarrow l_B$ and $l_A \rightarrow c$ crossovers and $\epsilon = 25$ for $p \rightarrow l_A$ and $l_B \rightarrow c$ crossovers. (b) The variation of the rate constants k_{c,l_A} and k_{c,l_B} with the temperature β for different chain lengths. (c) The variation of the rate constant $k_{l,p}$ with the temperature β for various chain lengths.

increasing N , as shown in Fig. 4(b) and 4(c), shows that the folding entropy loss rises with N . For loop formation, the scaling behavior of conformational entropy can be realized from the probability of nearest-neighbor contacts between two reactive sites, $\mathcal{P}[|\mathbf{r}_p - \mathbf{r}_q| \approx O(\sigma)] \sim N^{-\nu(3+\theta)} \mathcal{G}(L_1/L, L_2 + \ell_3/L)$.²⁵ Here θ is the correlation hole exponent which describes the short-distance spatial decay of the corresponding probability distribution due to the excluded volume interaction. Since $\Delta S/k_B = -\ln \mathcal{P} \sim \nu(3+\theta) \ln N$, it is reported²² that the dependence of the conformational entropy change on the length of a flexible chain can be expressed by

$$\Delta S = \alpha \ln N + G, \quad (11)$$

where the exponent $\alpha = \nu(3+\theta)$ is 1.98 for end-to-end loop, 2.16 for end-to-interior loop, and 2.48 for interior-to-interior loop. For a polymer with equal block lengths, the constant $G(L_1/L, L_2 + \ell_3/L) = G(L_1 + \ell_1/L, L_2/L)$ depends on the relative location between two reactive sites. As demonstrated in Fig. 5(a), all horizontal lines corresponding to different chain lengths fall into a single horizontal line in the plot of $\ln k_{c,l}^{-1}/N^{2.48}$ against β . This result indicates that the coil-to-loop folding kinetics is correctly depicted in our MC simulations.

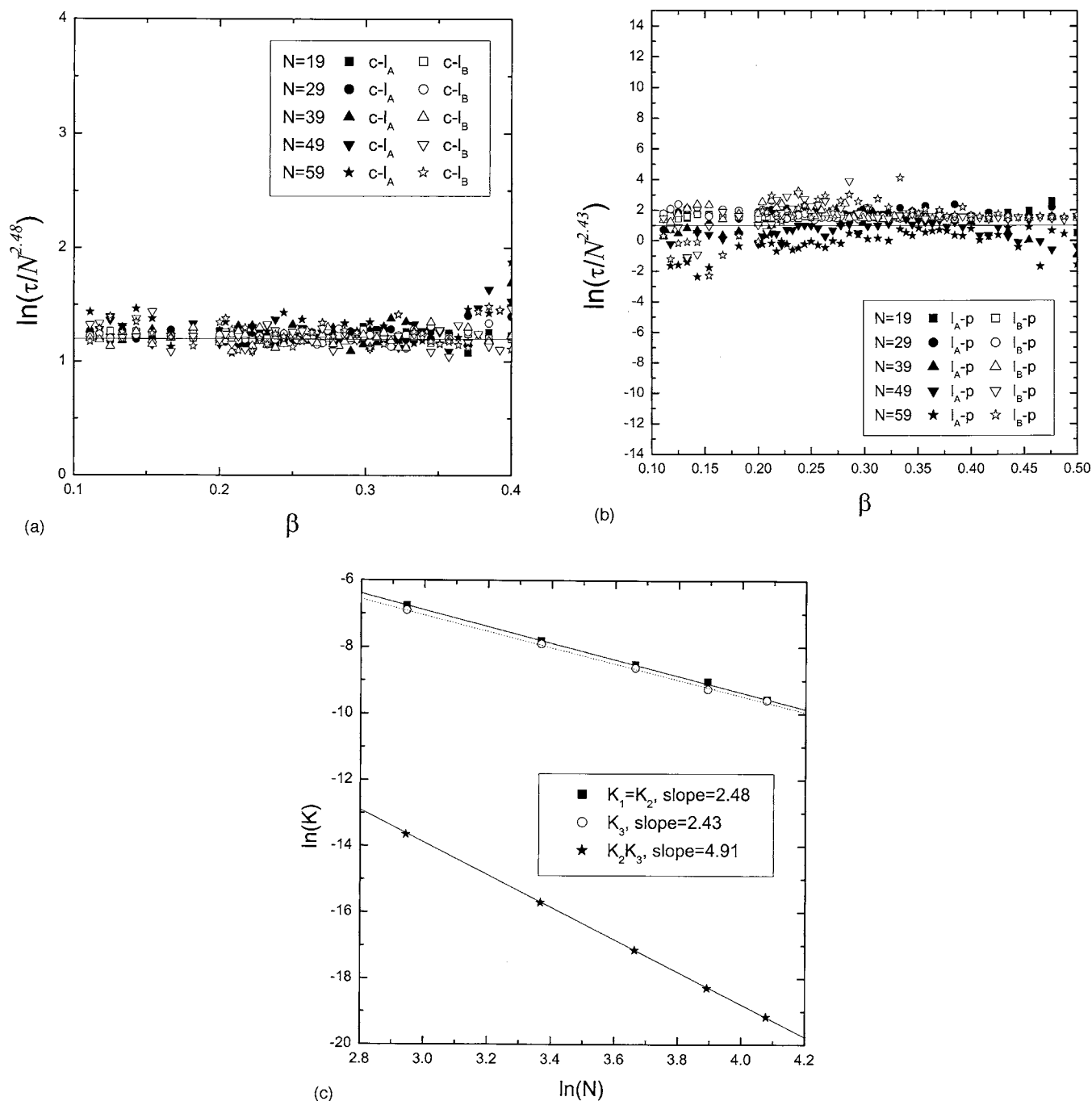


FIG. 5. (a) The scaling behavior of conformational entropy loss in the plot of $\ln k_{c,l}^{-1}/N^\alpha$ against β gives $\alpha=2.48$ for coil-to-loop crossover. (b) The scaling behavior of conformational entropy loss in the plot of $\ln k_{l,p}^{-1}/N^\alpha$ against β gives $\alpha=2.43$ for loop-to-pseudoknot crossover. (c) The scaling behavior of conformational entropy loss is displayed in the plot of $\ln K_i$ against $\ln N$. The slopes yield $\alpha=2.48$ for K_2 and $\alpha=2.43$ for K_3 .

The loop-to-pseudoknot folding kinetics involves another loop formation. It is natural to anticipate that the formation of the second loop may be disturbed by the existing loop. Nevertheless, we expect that the scaling expression, Eq. (10), is still valid but the exponent α and constant G change. In fact, after comparing the difference of $\ln k_{i,j}^{-1}$ between Fig. 4(b) for coil-to-loop and Fig. 4(c) for loop to pseudoknot, it is somewhat surprising to find that $\ln k_{l,p}^{-1}$ is only slightly greater than $\ln k_{c,l}^{-1}$ for various chain lengths. That is, for a flexible polymer with equal block lengths, the conformational entropy loss associated with coil-to-loop crossover is slightly smaller than that of loop-to-pseudoknot

crossover. Moreover, Fig. 5(b) shows that in the plot of $\ln k_{l,p}^{-1}/N^{2.43}$ against β , all horizontal lines for different chain lengths in Fig. 4(c) collapse into a single horizontal line. This consequence reveals that the exponent associated with loop-to-pseudoknot crossover is $\alpha \approx 2.43$, which is quite close to the exponent for coil-to-loop crossover. Nonetheless, as will be shown in Table II, the entropy loss for coil to loop A is generally different from that for loop B to pseudoknot.

The conformational entropy change can be extracted from either kinetic approach, Eq. (6), or thermodynamic approach, Eq. (4), based on results of Monte Carlo simulations. By fitting the probability curve with Eq. (3), one is able to

TABLE II. The influence of the distribution of loop lengths on the conformational entropy change for a polymer with total chain length $N=40$ and end block lengths $L_1=L_2=3$. (a) The case $\ell_1=\ell_2$, (b) the case $\ell_1=\ell_3$, and (c) the case $\ell_2=0$.

(a)	ℓ_1, ℓ_2, ℓ_3	$\Delta S_{c,A}/k_B$	$\Delta S_{c,B}/k_B$	$\Delta S_{i,B,p}/k_B$	$\Delta S_{c,p}/k_B$
	4,4,22	-9.55	-7.32	-9.57	-16.89
	6,6,18	-9.45	-8.05	-9.55	-17.60
	7,7,16	-8.42	-9.32	-8.62	-17.94
	8,8,14	-9.29	-8.68	-9.41	-18.09
	10,10,10	-9.03	-9.03	-9.31	-18.33
	13,13,4	-8.81	-9.53	-8.42	-17.95
	15,15,0	-8.62	-9.66	-8.18	-17.84
(b)	ℓ_1, ℓ_2, ℓ_3	$\Delta S_{c,A}/k_B$	$\Delta S_{c,i}/k_B$	$\Delta S_{i,B,p}/k_B$	$\Delta S_{c,p}/k_B$
	15,0,15	-8.52	-8.52	-9.11	-17.63
	13,4,13	-8.86	-8.86	-9.03	-17.89
	10,10,10	-9.03	-9.03	-9.31	-18.33
(c)	ℓ_1, ℓ_2, ℓ_3	$\Delta S_{c,i,A}/k_B$	$\Delta S_{c,i,B}/k_B$	$\Delta S_{i,B,p}/k_B$	$\Delta S_{c,p}/k_B$
	15,0,15	-8.52	-8.52	-9.11	-17.63
	10,0,20	-9.12	-7.78	-9.77	-17.55
	5,0,25	-9.43	-6.70	-9.85	-16.55

obtain all equilibrium constants, K_i . As shown in Fig. 5(c), the chain length dependence of the conformation entropy can be determined by the plot of $\ln K_i$ against $\ln N$. The slope for $\ln K_1=\ln K_2$ is -2.48 , which is consistent with the value of $-\alpha_1$ for coil to loop evaluated from the kinetic approach in Fig. 5(a). The slope for $\ln K_3$ is -2.43 , which agrees with the value of $-\alpha_2$ for loop to pseudoknot determined in Fig. 5(b). $\ln K_2K_3$ denotes the entropy loss from coil to pseudoknot $\Delta S_{c,p}$ and has a slope -4.91 , which is simply the sum of exponents associated with coil to loop and loop to pseudoknot. That is, the conformation entropy change for a flexible polymer can be expressed by

$$\Delta S_{p,c} = (\alpha_1 + \alpha_2) \ln N + G \left(\frac{L_i}{L}, \frac{\ell_i}{L} \right), \quad (12)$$

where the constant G varies with the relative ratios among end block lengths and loop lengths. When the total chain length is fixed, a biologically interesting question would be what the distribution of loop lengths would be to result in a minimum value of conformational entropy loss, i.e., G .

C. Asymmetric distribution in loop lengths

The pseudoknot formation is a result of the competition between binding energy gain and conformational entropy loss according to $\Delta G_p = -(\varepsilon_1 + \varepsilon_2) - T\Delta S_{c,p}$. For a given binding energy gain ($\varepsilon_1 + \varepsilon_2$), the smaller the entropy loss $\Delta S_{c,p}$, the larger the probability of pseudoknot formation P_p at a given temperature. It is therefore anticipated that for a given chain length N , a smaller value of G in Eq. (12) is preferred for pseudoknot formation. Recent examination of the PseudoBase database of pseudoknotted RNA structures reveals asymmetries in the loop lengths.¹⁹ Loop length ℓ_2 is often very short (172 of the 230 unique *ABAB* pseudoknots, or 75% have $\ell_2=0$; 195 of 230, or 85% have $\ell_2 \leq 1$). Such a structure might result in favorable coaxial helix stacking

interactions.¹⁹ Moreover, the asymmetry in loop lengths may lead to a decrease in conformational entropy loss. In the following study, we fix the total chain length and both end block lengths (L , L_1 , and L_2) and examine the distribution of loop lengths on the conformational entropy change.

Consider a flexible polymer with total chain length $N=40$. In order to observe the influence of the distribution of loop lengths $\{\ell_1, \ell_2, \ell_3\}$ on the conformational entropy change, we fix the two end block lengths, $L_1=L_2=3$. Since there are four reactive beads, the total loop length is $\ell_1 + \ell_2 + \ell_3 = 30$. Three cases (a) $\ell_1=\ell_2=\ell$, (b) $\ell_1=\ell_3$, and (c) $\ell_2=0$ will be examined. For the case (a) $\ell_1=\ell_2$, there are two combinations, $\{\ell, \ell, \ell_3\}$ and $\{\ell_1, \ell, \ell\}$. Although loops *A* and *B* are asymmetric in binding energy, the conformation entropy change for both situations should be the same because ΔS depends only on the distribution of block lengths. This equality is confirmed in our MC simulations. The asymmetry in loop lengths is therefore reflected by the ratio of ℓ_1 to ℓ_3 . As demonstrated in Table IIa, the entropy change is increased with increasing the symmetry, $\ell_1/\ell_3 \rightarrow 1$. The maximum entropy loss is obtained when $\ell_1=\ell_2=\ell_3$. That is, the greater the difference between ℓ_1 and ℓ_3 , the smaller the entropy loss.

In the second case (b) $\ell_1=\ell_3$, the asymmetry in loop lengths corresponds to the ratio of ℓ_1 to ℓ_2 . A similar trend to the first case is observed in Table IIb. The entropy loss declines with decreasing ℓ_2 and reaches the smallest value at $\ell_2=0$. The asymmetric characteristics can be further enhanced by considering the case (c) $\ell_2=0$ with $\ell_1 \neq \ell_3$. Now the asymmetry is manifested by the ratio of ℓ_1 to ℓ_3 . As displayed in Table IIc, the entropy loss is getting smaller with increasing the difference between ℓ_1 and ℓ_3 . The above results indicate that if the binding energy gain is the same, the probability of forming a pseudoknotted structure rises with increasing the asymmetric distribution in loop lengths. This consequence explains the experimental observation that the asymmetry in loop lengths is thermodynamically favorable in RNA pseudoknots.

V. CONCLUSION

Biomacromolecules such as RNA can fold into pseudoknot, consisting of non-nested double-stranded stems connected by single-stranded loops. The formation of such a topological structure involves the crossover from coil to loop and then loop to pseudoknot. By virtue of an athermal homopolymer of chain length N with two reactive pairs *AA* and *BB*, the folding kinetics and thermodynamics of *ABAB* pseudoknot formation are studied by off-lattice Monte Carlo simulations. The complication caused by the stem size is simplified by the binding between two reactive spheres. On the other hand, the excluded volume effects among nucleotides are fully included. Although our exceedingly simple model for pseudoknot formation is not the perfect representation of the RNA folding, it is sufficient to capture the essential feature of the physics for the prediction of the folding thermodynamics.

Since RNA folding kinetics is known to proceed stochastically via a succession of partially folded secondary struc-

tures in quasiequilibrium, kinetic Monte Carlo scheme is employed to follow the kinetics of pseudoknot formation. In our simulation, the binding energies are assumed to be $\varepsilon_1=10k_B T$ and $\varepsilon_2=25k_B T$, which correspond to about 3 and 6–7 base pairs. This is consistent with the frequencies of natural occurrence of stem lengths. Actually, our simple theory, which is confirmed by the Monte Carlo simulation indicates that the choice of ε_i has no influence on our result. Despite of the fact that there are two possible intermediate states for a simple pseudoknot polymer, loops *A* and *B*, the folding path prefers the one with more free energy reduction of coil-to-loop crossover. When the binding energy dominates the crossover, it is most probable that the polymer coil forms the loop with higher binding energy gain first. This consequence indicates that the denaturation of a pseudoknotted polymer starts from breaking the double-stranded stem with less base pairs.

For the approaches based on free energy minimization, the conformational entropy change associated with the formation of a pseudoknot plays an essential role in predicting the native structure of a RNA molecule. ΔS can be correctly extracted through equilibrium constants in probability curves or free energy barrier in kinetic rate constants, both of which are determined by Monte Carlo simulations. The entropy loss can be expressed by the scaling relation, $\Delta S = \alpha \ln N + G$, where the exponent $\alpha \approx 2.48$ for coil-to-loop crossover and $\alpha \approx 4.91$ for coil-to-pseudoknot crossover. The constant *G* associated with pseudoknot formation varies with the distribution of block lengths $\{L_1, \ell_1, \ell_2, \ell_3, L_2\}$ for a given *N*. For given end block lengths $\{L_1, L_2\}$, our simulation results reveal that as the three loop lengths are equal, i.e., $\ell_1 = \ell_2 = \ell_3$, the entropy loss (or *G*) is maximum. In other words, under such a structural characteristic, the total free energy reduction is smallest and the probability of pseudoknot formation is lowest. On the contrary, asymmetry in the distribution of loop lengths, such as $\{\ell_1, 0, \ell_3\}$, leads to smaller values of *G* and therefore favors the formation of pseudoknot thermodynamically. This consequence may elucidate the fact of loop length asymmetry in RNA pseudoknots reported recently.

For the free energy minimization methods, the well-known loop dependent energy rules assume additivity of loop free energy (entropy). We confirm that both loops in the pseudoknot have the same constant $\alpha \approx 2.48$ for the logarithmic increase of entropy with loop size. Note that the constant α is not 1.75 adopted in previous models. Nonetheless, our results also show that the conformational entropy constant *G* varies with the distribution of the three single-stranded loop

size $\{\ell_1, \ell_2, \ell_3\}$. This consequence reveal that the entropy constant may play an important role in studying pseudoknot thermodynamics. Our theoretical predictions, such as probability curve, rate constant, folding pathway, and pseudoknot entropy, can be tested by synthesizing a model RNA similar to our pseudoknotted homopolymer, such as the sequence $C \dots C(AA)C \dots C(AAAAA)C \dots C(UU)C \dots C(UUUUU)C \dots C$.

ACKNOWLEDGMENT

This research is supported by the National Council of Science of Taiwan.

- ¹C. W. A. Pleij, *Curr. Opin. Struct. Biol.* **4**, 337 (1994).
- ²D. S. McPheeters, G. D. Stormo, and L. Gold, *J. Mol. Biol.* **201**, 517 (1990).
- ³Z. Du, J. A. Holland, M. R. Hansen, D. P. Giedroc, and D. W. Hoffman, *J. Mol. Biol.* **271**, 463 (1997).
- ⁴P.-Y. Lai, Y.-J. Sheng, and H.-K. Tsao, *Phys. Rev. Lett.* **87**, 175503 (2001).
- ⁵A. Kabakçioğlu and A. L. Stella, *Phys. Rev. E* **70**, 011802 (2004).
- ⁶Y.-J. Sheng, C.-N. Wu, P.-Y. Lai, and H.-K. Tsao, *Macromolecules* **38**, 2959 (2005).
- ⁷A. Lucas and K. A. Dill, *J. Chem. Phys.* **119**, 2414 (2003).
- ⁸E. Rivas and S. R. Eddy, *J. Mol. Biol.* **285**, 2053 (1999).
- ⁹D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, *RNA* **299**, 911 (1999).
- ¹⁰M. Zuker, *Curr. Opin. Struct. Biol.* **10**, 303 (2000).
- ¹¹A. P. Gulyaev, F. H. D. van Batenburg, and C. W. A. Pleij, *RNA* **5**, 609 (1999).
- ¹²H. Isambert and E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6515 (2000).
- ¹³R. M. Dirks and N. A. Pierce, *J. Comput. Chem.* **24**, 1664 (2003).
- ¹⁴M. E. Fisher, *J. Chem. Phys.* **45**, 1469 (1966).
- ¹⁵A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isamber, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15310 (2003).
- ¹⁶G. Bonnet, O. Krichevsky, A. Libchaber, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3602 (1998).
- ¹⁷P. Leoni and C. Vanderzande, *Phys. Rev. E* **68**, 051904 (2003).
- ¹⁸M. Baiesi, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **91**, 198102 (2003).
- ¹⁹D. P. Aalberts and N. O. Hodas, *Nucleic Acids Res.* **33**, 2210 (2005).
- ²⁰Y.-J. Sheng, J. Z. Y. J. Chen, and H.-K. Tsao, *Macromolecules* **35**, 9624 (2002); Y.-J. Sheng, H.-J. Lin, J. Z. Y. Chen, and H.-K. Tsao, *J. Chem. Phys.* **118**, 8513 (2003).
- ²¹Y.-J. Sheng, J. Z. Y. Chen, and H.-K. Tsao, *Macromolecules* **36**, 5863 (2003).
- ²²Y.-J. Sheng, P.-H. Hsu, J. Z. Y. Chen, and H.-K. Tsao, *Macromolecules* **37**, 9257 (2004).
- ²³P. H. Verdier and W. H. Stockmayer, *J. Chem. Phys.* **36**, 227 (1962); P. H. Verdier, *ibid.* **59**, 6119 (1973); A. Baumgartner and K. Binder, *ibid.* **71**, 2541 (1979).
- ²⁴J. Z. Y. Chen, H.-K. Tsao, and Y.-J. Sheng, *Europhys. Lett.* **65**, 407 (2004); *Phys. Rev. E* **72**, 031804 (2005).
- ²⁵J. Des Cloizeaux and G. Jannink, *Polymers in Solution: Their Modelling and Structure* (Oxford University Press, Oxford, 1990), Chap. 13.