

# Folding thermodynamics of pseudoknotted chain conformations

Zoia Kopeikin and Shi-Jie Chen<sup>a)</sup>*Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211  
and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211*

(Received 18 January 2006; accepted 28 February 2006; published online 19 April 2006)

We develop a statistical mechanical framework for the folding thermodynamics of pseudoknotted structures. As applications of the theory, we investigate the folding stability and the free energy landscapes for both the thermal and the mechanical unfolding of pseudoknotted chains. For the mechanical unfolding process, we predict the force-extension curves, from which we can obtain the information about structural transitions in the unfolding process. In general, a pseudoknotted structure unfolds through multiple structural transitions. The interplay between the helix stems and the loops plays an important role in the folding stability of pseudoknots. For instance, variations in loop sizes can lead to the destabilization of some intermediate states and change the (equilibrium) folding pathways (e.g., two helix stems unfold either cooperatively or sequentially). In both thermal and mechanical unfolding, depending on the nucleotide sequence, misfolded intermediate states can emerge in the folding process. In addition, thermal and mechanical unfoldings often have different (equilibrium) pathways. For example, for certain sequences, the misfolded intermediates, which generally have longer tails, can fold, unfold, and refold again in the pulling process, which means that these intermediates can switch between two different average end-end extensions. © 2006 American Institute of Physics. [DOI: [10.1063/1.2188940](https://doi.org/10.1063/1.2188940)]

## I. INTRODUCTION

RNA pseudoknots are simple tertiary structures formed by the base pairing of nucleotides of a loop (hairpin, internal, bulge, or bifurcated) with nucleotides outside that loop. The simplest and most general is the hairpin (*H*)-type pseudoknot which includes only two loops and two stems: nucleotides of the hairpin loop base pair with nucleotides of the end of the chain. RNA pseudoknots are found to play a variety of important roles in cellular function: they are involved in translational regulation, viral replication, and some catalytic functions of ribozymes. In the present study, we develop a statistical mechanical thermodynamic model for RNA pseudoknots. Such a model may lead to a better understanding of RNA pseudoknot function.

Unlike secondary structures, which can be divided into nearly independent elements: loops, stems, and dangling segments, pseudoknot structures involve interactions which tie together secondary structure subunits (e.g., stems, loops) and make them strongly correlated. A result of the correlation between different structural subunits is the nonadditivity of chain entropy and free energy, i.e., the entropy and free energy of a pseudoknot cannot be calculated as a sum of entropies and free energies of each individual secondary structural subunits. For an *H* pseudoknot, which consists of two helix stems and two loops, the enthalpy can be estimated from the known energy parameters for helix stems,<sup>1-3</sup> but the calculation of the pseudoknot entropy requires a model that can account for the nonadditivity of chain entropy.

The estimation for the loop entropy for *H*-type

pseudoknot was first attempted by Gultyaev *et al.* (1999).<sup>4</sup> The authors used the Jacobson-Stockmayer approximation<sup>5</sup> for the loop entropy and fitted parameters from the stability data of known pseudoknots. Though the proposed thermodynamic parameters were shown to give much better approximation than the previous one,<sup>6</sup> the excluded volume effect (impossibility for two monomers to occupy the same site) has not been taken into account explicitly. Recently, a three-dimensional (3D) lattice-based statistical mechanical model<sup>7</sup> was developed to explicitly take into account the excluded volume effect for pseudoknots. However, the model disallows the formation of possible partially unfolded and misfolded intermediate states in the pseudoknot folding process.

In this paper, we develop a statistical mechanical model that can treat (a) the excluded volume effect, (b) the nonadditivity arising from the correlation between different structural subunits, and (c) the formation of partially folded and misfolded intermediates. Though our major focus here is on the *H* pseudoknot and the misfolded and the partially folded states, the methodology developed in this work is general and can be extended to treat more complex pseudoknotted structures. In addition, the method uses graphical representation for intrachain contacts and is thus independent of any specific chain representation. For illustrations, we use two-dimensional (2D) lattice chain conformations, where the excluded volume effect is accounted for by configuring the chain conformations as self-avoiding walks in a 2D lattice.

The paper is organized as follows. We first develop a statistical mechanical model for the conformational entropy of pseudoknotted structures. We then apply the model to study both the thermal and the mechanical folding free energy landscapes and folding thermodynamics for pseudoknotted structures. An advantage of the present theory

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: [ChenShi@missouri.edu](mailto:ChenShi@missouri.edu)

is its ability to treat the complete conformational ensemble for the pseudoknotted structures (and the secondary structures), including all the partially folded and misfolded states.

## II. GRAPH-THEORETIC APPROACH TO POLYMER THERMODYNAMICS

At the center of the statistical thermodynamics is the partition function  $Q(T)$ , defined as the weighted sum over all the possible conformational states,

$$Q(T) = \sum_{\text{conf}} e^{-E/k_B T}, \quad (1)$$

where  $k_B$  is Boltzmann's constant,  $T$  is the temperature, and  $E$  is the energy of an individual conformation. We use the graph-theoretic approach<sup>8-12</sup> to calculate the partition function. A polymer graph gives the intrachain contacts and is thus a convenient way to represent RNA structures. It consists of vertices which represent monomers, connected by straight line links as the (covalent) bonds between monomers. Monomers in spatial contact (base pairs in nucleic acids such as RNAs) are connected by curved links. There are three possible relationships between two contacts: nested, unrelated, and crossing linked.<sup>9</sup> Secondary structures involve only nested and unrelated contacts, while tertiary folds such as pseudoknots involve crossing links.

In terms of the polymer graph, the partition function can be calculated as a sum over all the possible graphs, which is computationally more efficient than over all the possible conformations,

$$Q(T) = \sum_{\text{graph}} \Omega e^{-E/k_B T}, \quad (2)$$

where  $\Omega$  is the number of chain conformations that satisfy the intrachain contacts defined in the graph. Since the energy  $E$  is determined by the intrachain contacts, a graph represents an equal-energy conformational ensemble. The key issue for the method is how to calculate the conformational count  $\Omega$  for the given graph. We start from the simple  $H$  pseudoknot and then generalize the method to treat more complex pseudoknots.

## III. CONFORMATIONAL ENTROPY OF PSEUDOKNOTS

The  $H$  pseudoknot consists of two stems and two loops, as shown in Fig. 1. For mathematical convenience, we cut off dangling tails, leaving only one bond of each tail to account for the excluded volume interactions between the tails and the rest of the structure. We call such tail-free structure "structure element." A full structure consists of two parts: structure element and the tails (in the 5' and 3' terminal regions of the nucleotide chain). Loops  $A$  and  $B$  share the common single-stranded (interfacial) chain segment and are also constrained by the helical stems, therefore, the conformations of loops  $A$  and  $B$  are strongly correlated with each other. The basic idea in our pseudoknot conformational entropy theory is to divide a pseudoknot into three components, namely, the enlarged interface (shown as solid lines in Fig. 1), and the two noninterfacial free loop segments  $F_A$  and  $F_B$  (shown as dashed lines in Fig. 1) of lengths  $f_A$  and  $f_B$ , re-

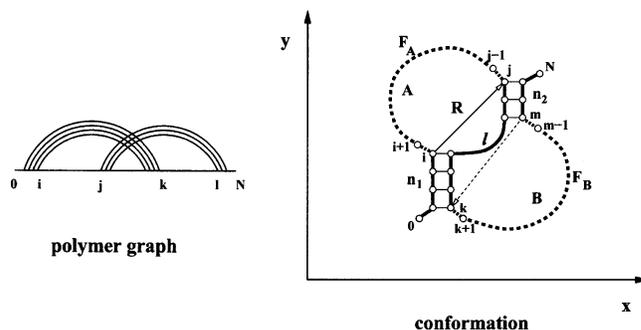


FIG. 1. The polymer graph and the structure of an  $H$  pseudoknot. The pseudoknot is divided in our theory into enlarged interface (bold lines) and two free loop segments  $F_A$  and  $F_B$  (dashed lines). The coordinate system is chosen so that the  $y$  axis is along the first pseudoknot stem. The number of free loop segment conformations is a function of the length of the free loop segment and the end-end vector  $\mathbf{R}$ .

spectively. The enlarged interface includes two stems of lengths  $n_1$  and  $n_2$ , the single-stranded interfacial segment of length  $l$ , and the tail monomers 0 and  $N$  in Fig. 1.

A great challenge in the entropy calculation is how to treat the excluded volume effect. According to the decomposition of the pseudoknot structure, we classify the excluded volume interactions into two types: (a) between the interfacial monomers (i.e., monomers in the enlarged interface) and (b) between the free loop monomers (i.e., monomers in the free loop segments) and between the free loop monomers and the interfacial monomers. We treat the former (type a) excluded volume effect by considering all the viable self-avoiding enlarged interface conformations (denoted as  $\mathbf{I}$ ) that (i) do not make self-contacts other than those specified by the graph and (ii) allow viable positioning of monomers  $i+1$ ,  $j-1$ ,  $k+1$ , and  $m-1$  without causing additional contacts. In terms of  $\mathbf{I}$ , we compute the number of conformations  $\Omega_P$  of the pseudoknot as a sum over all the possible viable conformations  $\mathbf{I}$  of the enlarged interface,

$$\Omega_P = \sum_{\mathbf{I}} \Omega_f(\mathbf{I}, f_A) \cdot \Omega_f(\mathbf{I}, f_B), \quad (3)$$

where  $\Omega_f(\mathbf{I}, f_A)$  and  $\Omega_f(\mathbf{I}, f_B)$  are the numbers of conformations of the free loop segments  $F_A$  and  $F_B$  for a given  $\mathbf{I}$ , respectively. The later (type b) excluded volume effect plays a crucial role and should be taken into account in the computation of  $\Omega_f(\mathbf{I}, f_A)$  and  $\Omega_f(\mathbf{I}, f_B)$ .

In Eq. (3), we neglect the volume exclusion between the free loop segments  $F_A$  and  $F_B$ . This is because, first, the free loop segments are spatially separated by the enlarged interface and, second, the steric hindrance between each free loop segment and the interface is accounted for in the calculation of  $\Omega_f$  (see next section). For very large loops and long free loop segments, the volume exclusion between  $F_A$  and  $F_B$  may become important. However, in that case, the chain entropy is large, so the error caused by neglecting the  $F_A$ - $F_B$  volume exclusion is relatively small as compared with the (large) chain entropy.

If one has a table for  $\Omega_f(\mathbf{I}, f_A)$  and  $\Omega_f(\mathbf{I}, f_B)$  for all the possible free loop segment lengths  $f_A$  and  $f_B$  and enlarged interface conformations  $\mathbf{I}$ , the computation of  $\Omega_P$  for the given graph (pseudoknot) from Eq. (3) would be efficient

and straightforward. However, the number of possible parameter sets  $(f_A, f_B, \mathbf{I})$ , grows exponentially with the length of the interface, which makes the tabulation of  $\Omega_f$  for all the possible parameter sets practically impossible. In the next section, we develop a method to approximate the conformational count  $\Omega_f$  for the free loop segments. The derived expression for  $\Omega_f$  would be mathematically convenient and computationally efficient.

### A. Number of conformations of a free loop segment

Because of the obviously equal roles of loops  $A$  and  $B$  in Fig. 1, instead of calculating  $\Omega_f$  for the two loops separately, we focus on the calculation for one of the loops, namely, loop  $A$ . In what follows, we derive an approximate expression for the number of conformation  $\Omega_f$  for the free loop segment  $F_A$  in loop  $A$ . The derived result for loop  $A$  would be equally applicable to loop  $B$ .

Strictly speaking, the number of conformations  $\Omega_f = \Omega_f(\mathbf{I}, f_A)$  for the free loop segment  $F_A$  (in Fig. 1) is a function of the enlarged interface conformation  $\mathbf{I}$ , which is dependent on the interface length  $l$  and the length of the two helical stems ( $n_1$  and  $n_2$ ), etc. However, as an approximation, we replace the  $\mathbf{I}$  dependence by the end-end vector  $\mathbf{R}$  dependence; see the vector from  $i$  to  $j$  in Fig. 1. The end-end vector for  $F_B$  is shown as the dashed line in Fig. 1. We further define a two-dimensional Cartesian coordinate system by choosing the  $y$  axis along stem 1 (directed upward), as shown in Fig. 1; vector  $\mathbf{R}$  can be described through the components:  $\mathbf{R}=(x, y)$ , where  $x=x_j-x_i$  and  $y=y_j-y_i$  in Fig. 1. In terms of the components of  $\mathbf{R}$ , we reduce the  $\Omega_f$  function from a large parameter space (for the interfacial chain conformations  $\mathbf{I}$ ) to a much smaller two-variable parameter space [for  $\mathbf{R}=(x, y)$ ],

$$\Omega_f(\mathbf{I}, f) \approx \Omega_f(\mathbf{R}, f) \rightarrow \Omega_f(x, y, f).$$

Since the information about  $\mathbf{I}$  is now embedded in  $\mathbf{R}$ , for an  $f_A$ -mer free loop segment  $F_A$  with fixed end-end vector  $\mathbf{R}=(x, y)$ , we compute the conformational count  $\Omega_f(\mathbf{R}, f_A)$  as an average over all the possible enlarged interface conformations  $\mathbf{I}$  with the fixed end-end vector  $\mathbf{R}$ ,

$$\Omega_f(\mathbf{R}, f_A) \approx \frac{\sum_{\mathbf{I}} \Omega_f(\mathbf{I}, f_A)}{\Omega_{\mathbf{I}}},$$

where  $\Omega_{\mathbf{I}}$  is the number of all the viable enlarged interface conformations with the given  $\mathbf{R}=(x, y)$ , and  $\sum_{\mathbf{I}}$  is the sum over all these conformations. Physically, the calculated  $\Omega_f(x, y, f_A)$  function is the average conformational count of  $F_A$  for each given enlarged interface conformation  $\mathbf{I}$ .

In fact, the first helix stem length  $n_1$  affects the conformational count of  $F_A$  only through weak excluded volume interactions mainly located around the junction between the helix and  $F_A$ . For the helix stem length  $n_1 \geq 2$  base stacks, the conformational count for  $F_A$  is nearly independent of  $n_1$ . In our calculation, we choose  $n_1=3$  base stacks and fix the orientation of helix stem 1 to be upright (in the  $y$  direction). We then enumerate all the possible conformations of the enlarged interface for all the possible values of  $l$  and  $n_2$  such that  $1 \leq l+n_2 \leq 10$ . For each enlarged interface conformation

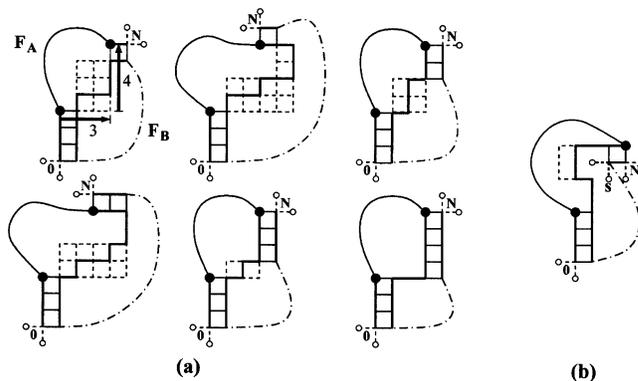


FIG. 2. (a) All the viable conformations of possible enlarged interfaces that give end-end vector  $\mathbf{R}=(x, y)=(3, 4)$  for the free loop segment  $F_A$ . The position and length of the first stem are fixed. Big black circles denote the end monomers of the free loop segment  $F_A$ . Part of the 2D lattice which can be covered by the viable conformations of the interfacial segment for possible length and position of the second stem is shown with dashed lines. The dash-dotted line denotes the  $F_B$  segment. Two possible positions for 0 and  $N$  are shown for each interfacial conformation. (b) An example of the conformation of the enlarged interface which is not viable because all the possible positions of the monomer  $s$  would make an additional contact with the interfacial segment.

$\mathbf{I}$  and the given free loop segment length  $f_A$  ( $f_A \leq 21$ ), we calculated the number of free loop segment conformations  $\Omega_f(\mathbf{I}, f_A)$  by means of exact computer enumerations of self-avoiding walks on 2D lattice.

As an illustration, in Fig. 2(a), we show all the viable conformations of the enlarged interface with fixed  $n_1=3$  (base stacks) and end-end vector  $\mathbf{R}=(x, y)=(3, 4)$ . For the given end-end vector  $\mathbf{R}$ , the length  $n_2$  of stem 2 has only four possible viable values:  $n_2=1, 2, 3, 4$ , and the length of the single-stranded interfacial segment is restricted by  $l \leq 10 - n_2$ . The possible interface conformations are represented as self-avoiding random walks on the dashed grid (with one of the conformations drawn as solid line). Figure 2(b) shows how the volume exclusion between the free loop segment and the interface determines the viability of conformations. The depicted conformation of the enlarged interface in Fig. 2(b) is not viable because it is impossible to position monomer  $s$  in the square lattice without making additional contacts with the single-stranded interfacial segment.

### B. Number of conformations of pseudoknots

#### 1. Pseudoknot with two stems

With the reduced function  $\Omega_f(\mathbf{R}, f)$  for free loop segments  $F_A$  and  $F_B$ , we obtain the conformational count of the pseudoknot from Eq. (3) by replacing  $\Omega_f(\mathbf{I}, f)$  by  $\Omega_f(\mathbf{R}, f)$ ,

$$\Omega_P \approx \sum_{\mathbf{I}} \Omega_f(\mathbf{R}_A, f_A) \cdot \Omega_f(\mathbf{R}_B, f_B),$$

where  $\mathbf{R}_A$  and  $\mathbf{R}_B$  are the end-end vectors of the free loop segments  $F_A$  and  $F_B$ , respectively, for a given conformation  $\mathbf{I}$  of the enlarged interface. As shown in Fig. 1,  $\mathbf{R}_A$  is the vector  $i \rightarrow j$  and  $\mathbf{R}_B$  is the vector  $m \rightarrow k$ .

The method can be generalized to treat more complicated pseudoknots. Here we demonstrate how to extend the method to treat pseudoknots with an internal loop in the stems [Fig. 3(a)]. Such ‘‘pseudoknot’’ conformations are im-

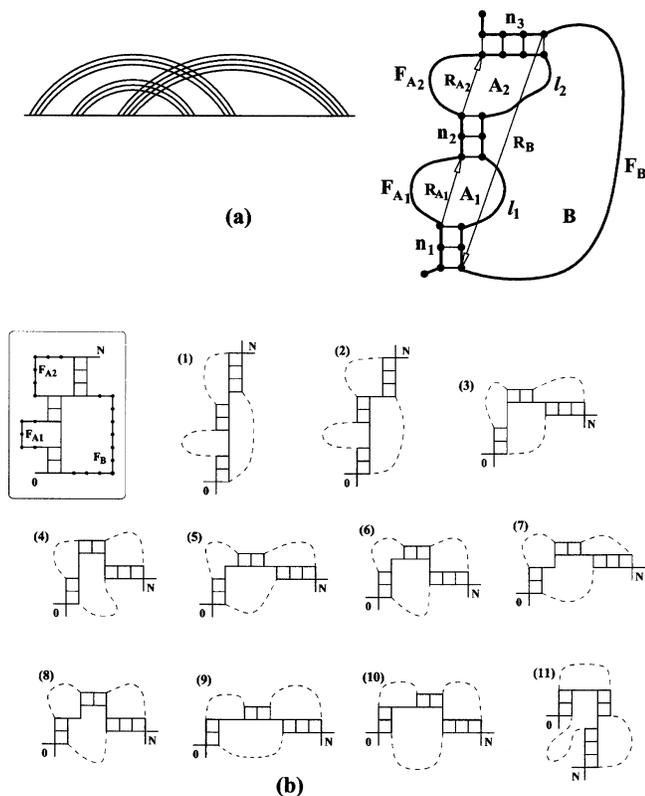


FIG. 3. (a) The polymer graph and structure of a “pseudoknot” with an internal loop ( $A_1$ ) formed in a helix stem.  $R_{A_1}$ ,  $R_{A_2}$ , and  $R_B$  shown in the figure are the end-end vectors of the free loop segments  $F_{A_1}$ ,  $F_{A_2}$  and  $F_B$ , respectively. (b) The structure of the pseudoknot with three stems chosen to illustrate the method and the viable conformations of the enlarged interface. Two positions are shown for each of the end monomers 0 and  $N$ .

portant because they may emerge as partially folded or misfolded intermediates in the pseudoknot folding process.

## 2. Generalized pseudoknots with three stems

As shown in Fig. 3, the enlarged interface for such a structure consists of three stems ( $n_1, n_2, n_3$ ), two single-stranded interfacial segments ( $I_1, I_2$ ), and two end monomers (0,  $N$ ). For a given conformation  $\mathbf{I}$  of the enlarged interface, the end-end vectors  $R_{A_1}$ ,  $R_{A_2}$ , and  $R_B$  of the free loop segments  $F_{A_1}$ ,  $F_{A_2}$ , and  $F_B$ , respectively, can be unambiguously

determined. The conformational count of the pseudoknot in Fig. 3 can be calculated from the following sum over all the possible conformations  $\mathbf{I}$  of the enlarged interface:

$$\Omega_P \approx \sum_{\mathbf{I}} \Omega_f(\mathbf{R}_{A_1}, f_{A_1}) \cdot \Omega_f(\mathbf{R}_{A_2}, f_{A_2}) \cdot \Omega_f(\mathbf{R}_B, f_B). \quad (4)$$

## 3. Illustrative calculation for the number of conformations of pseudoknots with three stems

In Fig. 3(b), we show the structure of a pseudoknot with an internal loop formed in one of the helix stems so the pseudoknot is partially folded (unfolded) and contains three helix stems. Also shown in the figure are all the 11 viable conformations of the enlarged interface numbered from 1 to 11 (excluding the end monomers 0 and  $N$ ). Since monomers 0 and  $N$  each can have two possible positions, from Eq. (4), we calculate the number of pseudoknot conformations as the following:

$$\Omega_P \approx (2 \times 2) \sum_{\mathbf{K}=1}^{11} \Omega_f(\mathbf{R}_{A_1}, f_{A_1}) \cdot \Omega_f(\mathbf{R}_{A_2}, f_{A_2}) \cdot \Omega_f(\mathbf{R}_B, f_B), \quad (5)$$

where  $\mathbf{K}$  denotes an interface conformation shown in Fig. 3(b). The lengths of the free loop segments for the given pseudoknot are  $f_{A_1}=5$  for  $F_{A_1}$ ,  $f_{A_2}=6$  for  $F_{A_2}$ , and  $f_B=11$  for  $F_B$ . The details of calculations are given in Table I, where we show, for each interfacial conformation  $\mathbf{K}$ , the coordinates ( $x, y$ ) of the end-end vectors and the approximate numbers of conformations  $\Omega_f$  for each free loop segment ( $F_{A_1}$ ,  $F_{A_2}$ , and  $F_B$ ).

As shown in Table I, Eq. (5) gives  $\Omega_P^{\text{appr}} = 4 \times 791.1 = 3164.4$ , which is close to the result from the exact computer enumeration:  $\Omega_P^{\text{exact}} = 2799$ .

## IV. THERMAL UNFOLDING OF PSEUDOKNOTS

Central to the folding thermodynamics is the partition function. The calculation of partition function [Eq. (2)] for a given nucleotide sequence requires the enumeration of all the possible polymer graphs and the counting of conformations accessible to each graph. In the present study, we consider the complete ensemble of pseudoknotted structures as well

TABLE I. Illustrative calculations for the number of conformations for a partially folded pseudoknot shown in Fig. 3(b).

$\mathbf{K}$	$(x_{A_1}, y_{A_1})$	$\Omega_f^{A_1}$	$(x_{A_2}, y_{A_2})$	$\Omega_f^{A_2}$	$(x_B, y_B)$	$\Omega_f^B$	$\Omega_f^{A_1} \cdot \Omega_f^{A_2} \cdot \Omega_f^B$
1	(0,2)	1.0	(1,4)	2.6	(1, 7)	23.2	60.3
2	(0,2)	1.0	(2,3)	3.0	(2,6)	55.5	166.5
3	(1,3)	1.8	(1,4)	2.6	(-3, 3)	21.6	101.1
4	(1,3)	1.8	(2,3)	3.0	(-2, 2)	8.0	43.2
5	(2,2)	1.1	(1,4)	2.6	(-2, 4)	32.1	91.8
6	(2,2)	1.1	(2,3)	3.0	(-1, 3)	17.0	56.1
7	(2,2)	1.1	(1,4)	2.6	(-2, 4)	32.1	91.8
8	(2,2)	1.1	(2,3)	3.0	(-1, 3)	17.0	56.1
9	(3,1)	0.5	(1,4)	2.6	(-1, 5)	52.5	68.2
10	(3,1)	0.5	(2,3)	3.0	(0,4)	37.3	55.9
11	(4,0)	0.0	(1,4)	2.6	(-1, 1)	0.0	0.0
							Sum=791.1

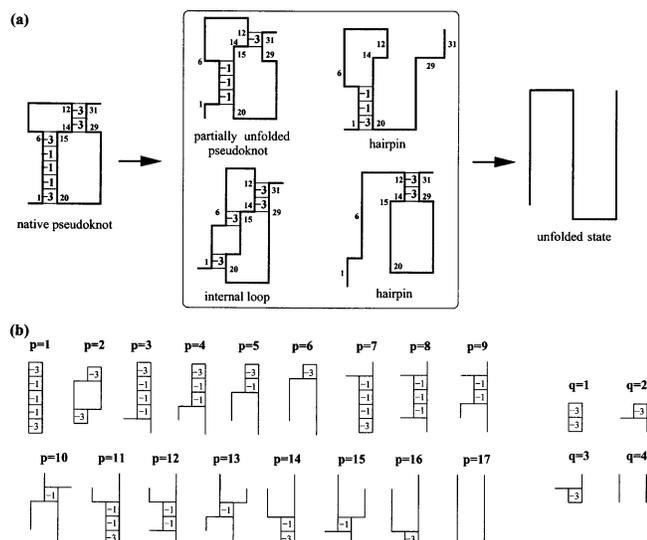


FIG. 4. (a) The 33-mer pseudoknot-forming chain is used to illustrate the calculations for the density of states and partition function. The numbers in boxes (stacks) denote the energies of the corresponding (native) base stack. The four types of the representative partially unfolded conformations considered in the partition function calculation are shown. (b) The partially unzipped states of the (two) helix stems are labeled with parameters  $p$  and  $q$ , respectively. As a result, each state (folded, partially folded, and unfolded) can be represented by the parameter pair  $(p, q)$ .

as the secondary structures. For the secondary structure partition function, we use a previously developed statistical mechanical theory.<sup>10,11</sup>

For the pseudoknotted conformations, in the preceding sections, we ignored the full tails and took into account only the first monomer of each tail closest to the pseudoknot structure element to represent the volume exclusion effect. We now add the full tails back to the calculation.

To take into account tails, we have used precalculated<sup>11</sup> tables for the numbers of tail conformations  $\Omega_T(t)$  for tail length  $t \leq 24$  and the following fitted formula for longer tails:<sup>11</sup>

$$\ln \Omega_T(t) \approx -2.62208 + 0.83927t + 0.30984 \ln(t).$$

This formula has also been used to obtain the number of conformations of the open (fully unfolded) chain. With the conformational count of the tails, we can calculate the number of conformations of the pseudoknot with tails as the product of the number of conformations of the pseudoknot structure element  $\Omega_P$  and the numbers of conformations of tails:  $\Omega = \Omega_P \Omega_T(t_1) \Omega_T(t_2)$ , where  $t_1$  and  $t_2$  are lengths of the two tails at the 5' and the 3' terminals, respectively.

### A. The Go-model pseudoknots

We consider a 33-mer chain with a fully folded native state shown in Fig. 4(a). As a simplified (Go-type) model, we assume that no other contacts besides those (native contacts) depicted in Fig. 4(a) can form. The interaction energy of each base pair stack is assigned to be  $-3$  or  $-1$  as shown in the figure. The energy of an isolated (unstacked) base pair is assumed to be 0.

Since in the Go model, only the native contacts can form or disrupt in the folding/unfolding process, the conforma-

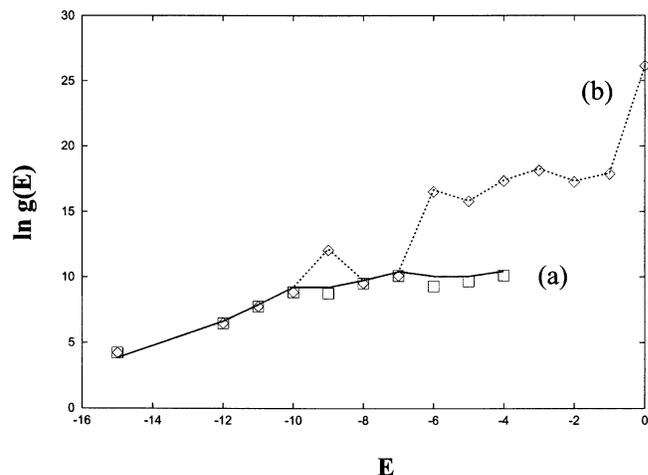


FIG. 5. Test of the theory (line) against exact computer enumeration (symbols) for the density of states (a) for all the pseudoknotted conformations and (b) for the complete pseudoknot/hairpin/open conformational ensemble for the 33-mer chain shown in Fig. 4(a).

tional ensemble can be generated from the different ways to break the native contacts. Examples of partially unfolded states are also shown in Fig. 4(a). The unzipping of either helix stem can occur at the top, bottom, or internal base pair of the stem. The partially unfolded states of the pseudoknot can be conveniently represented by two parameters:  $p$  denotes the state of stem 1 and  $q$  of stem 2. The partially unfolded states of a stem depend on the stem length and can be enumerated. The possible states of stems of our native pseudoknot (five- and two-base stack) are shown in Fig. 4(b). In this way, each possible pseudoknot state can be described by a pair of numbers  $(p, q)$ , which defines the set of contacts. For example,  $(p, q) = (1, 1)$  is the native pseudoknot in Fig. 4(a),  $(2, 4)$  is a hairpin with an internal loop and  $(2, 3)$  is a pseudoknot with an internal loop in stem 1.

The parameters  $p$  and  $q$  are used as the labels for the states of the stems. Each  $(p, q)$  pair unambiguously defines a pseudoknotted structure. The  $p$ 's and  $q$ 's shown in Fig. 4 are exhaustive for the short stems shown in the figure. As a caveat, we note that for illustrative purposes, we here use the two-dimensional lattice model, which excludes some conformations due to the lattice constraint (e.g., a 6-mer loop is not possible in a two-dimensional square lattice). In addition, in this section, in order to focus on the stem-loop interplay, we use the Go model, which disallows the formation of the misfolded states. For longer stems, more complex multiple internal loops can be formed. In the next section, we will go beyond the Go model by using the complete conformational ensemble, including all the possible misfolded structures.

To test the accuracy of the theory, we compute the density of states for all the possible pseudoknots (including all the partially unfolded states) and for the complete (pseudoknot/hairpin/open chain) conformational ensemble (including all the possible partially unfolded states) using both the theory developed here and the exact computer enumeration. Figure 5 shows the test results. We find good accuracy of the theoretical prediction.

To study the folding thermodynamics for the model pseudoknot in Fig. 4(a), we calculate the free energy land-

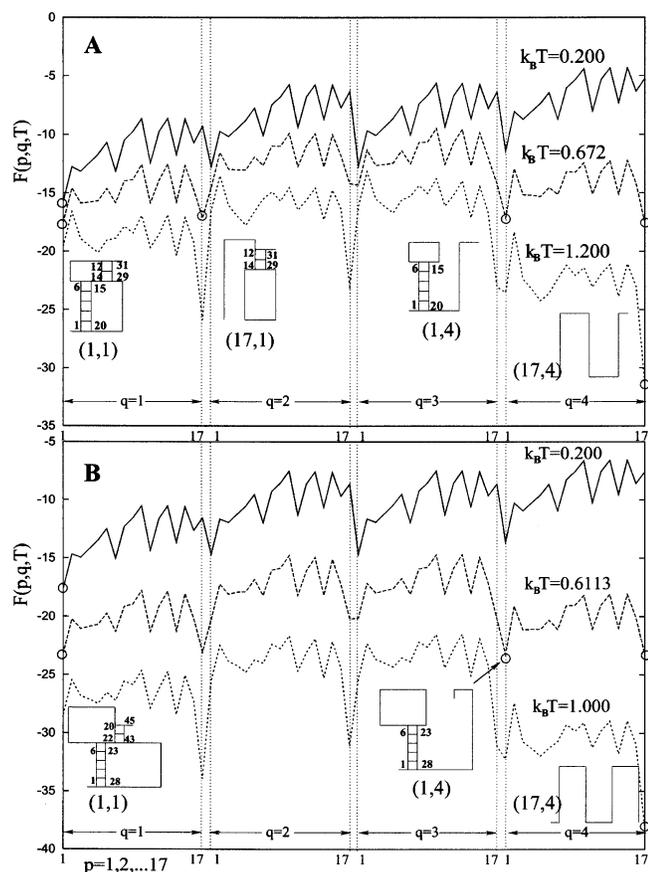


FIG. 6. The free energy  $F(p,q,T)$  as a function of structure denoted by parameter pair  $(p,q)$ . Free energy minima, i.e., stable states for a given temperature, are circled, and the representative conformations shown. (A) For the pseudoknot with smaller loops, the two hairpins (17, 1) and (1, 4), each formed through the disruption of a native helix stem, emerge as stable intermediate states. (B) For the pseudoknot with larger loops, the transition is three state: the native state, open chain, and hairpin with the longer helix stem, which are equally populated at  $k_B T=0.6113$ .

scape  $F(p,q)$ , which is the free energy of the macrostate for all the possible secondary structures and pseudoknots described by the conformational state  $(p,q)$ .  $F(p,q)$  is computed from the following equation:

$$F(p,q) = E(p,q) - k_B T \ln \Omega(p,q),$$

where  $E(p,q)$  and  $\Omega(p,q)$  are the energy (sum of the energies of the stacks) and the number of conformations of the macrostate described by  $(p,q)$ . The free energy minima correspond to stable states.  $F(p,q)$  is temperature  $T$  dependent, and temperature change causes the change in the free energy landscape  $F(p,q)$  and the transitions between different stable states.

Figure 6(a) shows the free energy landscape  $F(p,q)$  for the model pseudoknot in Fig. 4(a). The landscape shows single pronounced minimum at  $(p,q)=(1,1)$  [= the native state shown in Fig. 4(a)] at low temperature and (17,4) (= fully unfolded state) at high temperature. At  $k_B T=0.672$ , the landscape shows that the transition between the native state and the unfolded state involves two hairpin conformations as the intermediate states: (17,1) and (1,4). Each hairpin intermediate is formed through the disruption of a helix stem of the native pseudoknot.

In order to examine the competition between the helix and loop stability, we further calculate the free energy landscape for pseudoknots with different sizes of loops. For larger loops, as shown in Fig. 6(b), the free energy landscape reveals three-state transitions, where the native state, the open chain, and the hairpin with the longer stem are equally populated at  $k_B T=0.6113$ . The larger loop would destabilize the folded state. As a result, state (17,1), which emerges as a folding intermediate for pseudoknot with smaller loops, is now absent. This is because the two-stack short helix stem in the (17,1) state is not stable enough to compete with the destabilizing larger loop. In contrast, the state (1,4) emerges as an intermediate because the five-stack long helix stem is sufficient enough to compete with (the larger) loop.

## B. Pseudoknot folding with misfolded states

The formation of non-native contacts is not allowed in the above Go-model pseudoknots and therefore many structures, namely, the misfolded states, are excluded from consideration. In this section, we treat complete conformational ensemble, including all the possible misfolded conformations. We choose two (30- and 34-nt) pseudoknot-forming nucleotide sequences. We allow the formation of all the possible  $A-U$  and  $C-G$  base pairs. The energy of a base stack is equal to  $-1$  for a stack formed by two  $A-U$  base pairs,  $-3$  for a stack formed by one  $A-U$  and one  $C-G$  base pair, and  $-4$  for a stack formed by two  $C-G$  pairs.

We enumerate all the possible secondary structures and pseudoknotted structures through the enumeration of all the possible polymer graphs. For each graph, we compute the energy  $E$  and the number of accessible conformations  $\Omega$ . In Figs. 7 and 8 we show the free energy landscapes ( $F=E - k_B T \ln \Omega$ ) for the 30-nt and the 34-nt sequences, respectively. The free energy landscapes show contrasting folding thermodynamics for the two sequences. The sequence in Fig. 7 unfolds through the sequential disruption of the two helix stems. The landscape at  $k_B T=1.0$  shows the emergence of an intermediate state (= minimum in the free energy landscape) which is formed through the breaking of the less stable two-stack helix in the native pseudoknot. In contrast, the sequence in Fig. 8 unfolds through the formation of a misfolded state I, shown as the minimum in the free energy landscape at  $k_B T=0.9$ . The misfolded intermediate (hairpin I in Fig. 8) is formed through a complete rearrangement of the base pairs from the native pseudoknot.

## V. MECHANICAL UNFOLDING OF PSEUDOKNOTS

In the mechanical unfolding experiments, a stretching force in the piconewton (pN) range generated by atomic force microscopy or laser or magnetic tweezers is applied to pull a biomolecule to unfold the molecule.<sup>13-15</sup> Such experiments have been conducted on proteins,<sup>16-18</sup> DNA,<sup>19</sup> and RNA.<sup>20-23</sup> The advantage of the single-molecule experiments is the possibility to follow the individual folding-unfolding trajectory (which is difficult in bulk studies where multiple species and multiple folding pathways are present) and to stretch the molecule along the well-defined coordinates.

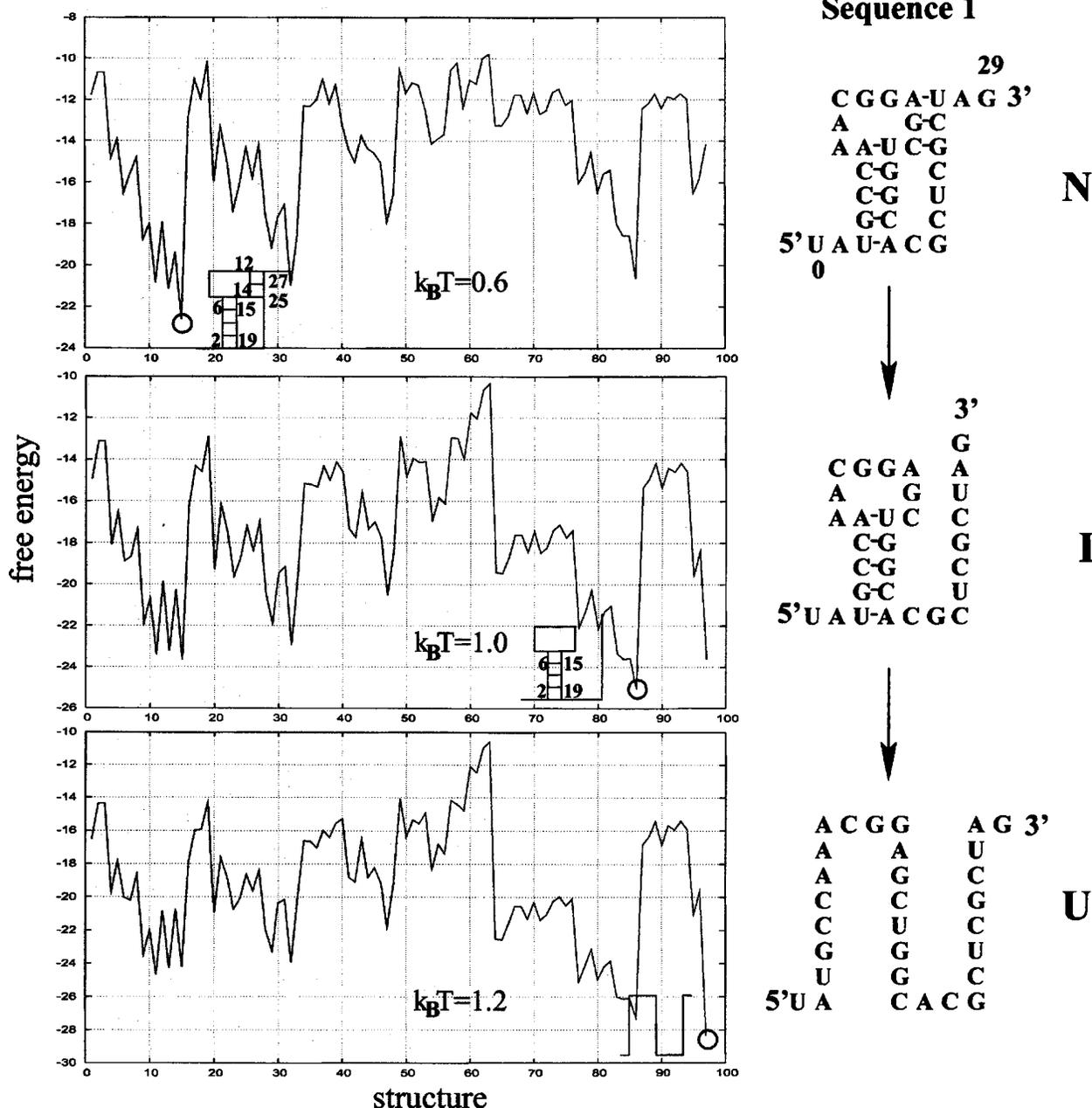


FIG. 7. The free energy as a function of structure for a 30-mer nucleotide sequence. The unfolding is a two-step process with the (less stable) shorter helix stem disrupted first ( $N \rightarrow I$ ) followed by the breaking of the longer (more stable) helix stem ( $I \rightarrow U$ ).

The mechanical unfolding experiments have inspired the theoretical studies on the force induced folding/unfolding of biomolecules.<sup>24-27</sup> One of the important features in the force induced unfolding is the dependence on the choice of the statistical ensemble, namely, which variables are held constant and which are allowed to fluctuate in the unfolding process. In the experiments where the molecule is stretched by force, the following two types of statistical ensembles are generally considered.

(a) Constant distance ensemble (isometric): The molecule's end-end distance is held constant, and the fluctuation of the force is recorded in experiment. Such an experiment can be performed, for example, in the following way. One end of the molecule can be attached to a rigid support, and the other end is attached to the bead held in optical trap. The

feedback loop controls the position of the trapped bead with respect to the other end of the molecule and cancels its fluctuations by moving the trap center. The fluctuating force, which is proportional to the displacement of the bead in the optical trap from its equilibrium position, as a function of time can be determined from the movement of the optical trap. In the equilibrium unfolding process, the molecule's end-end distance (extension)  $D$  is changed slowly, and the mean force averaged over the appropriate time period  $\bar{F}$  as a function of  $D$  is plotted as the force-extension curve.

(b) Constant force ensemble (isotensional): The force is held constant, and the molecule's end-end distance fluctuates. In such experiments, the feedback loop makes the optical trap move in such a way that the bead position with

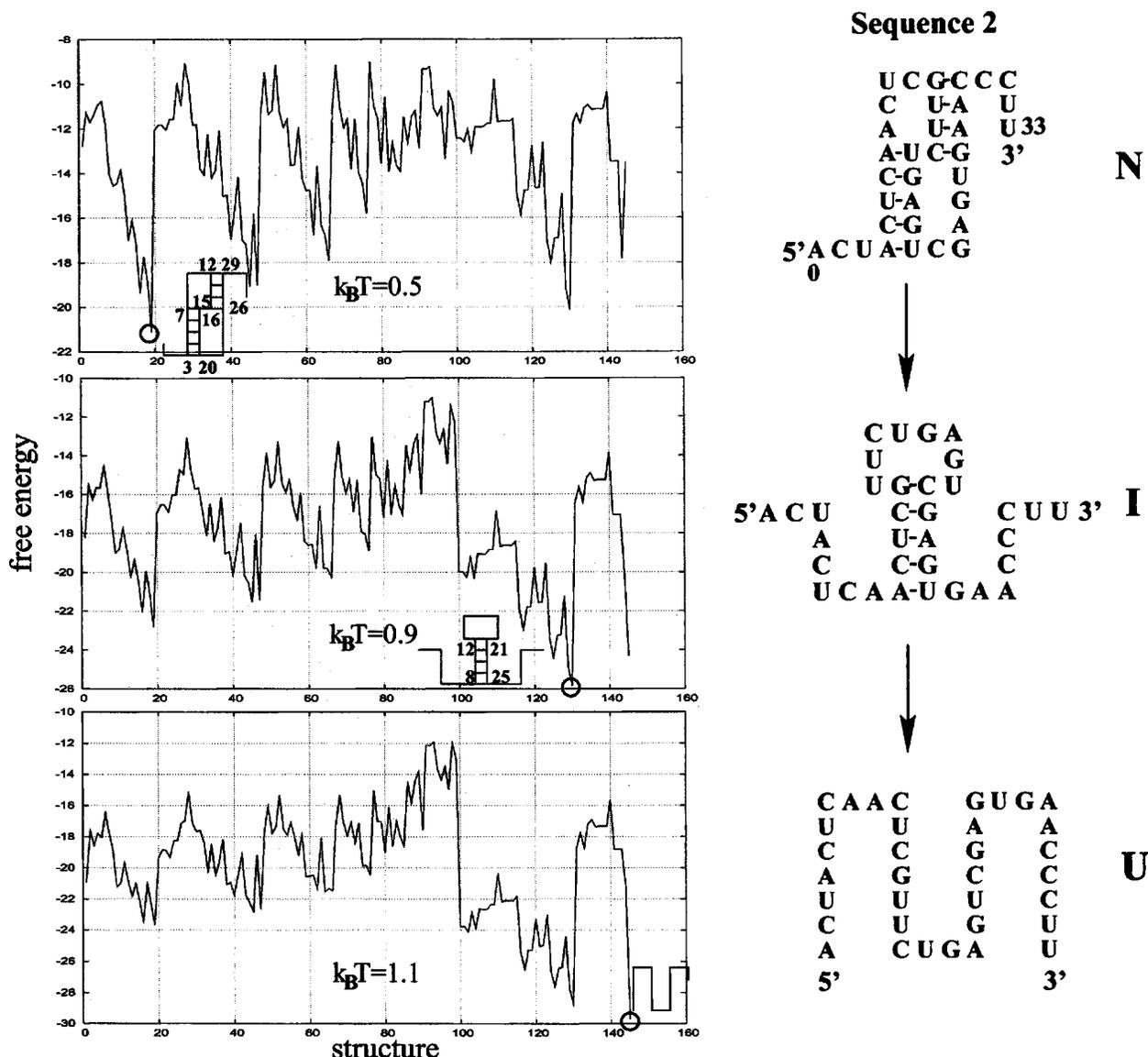


FIG. 8. Free energy as a function of structure. The melting process for the 34-mer nucleotide sequence involves a misfolded hairpin (*I*) as an intermediate state.

respect to the center of the optical trap is fixed (which means that the fixed force is acting at the bead). The fluctuating end-end distance is then averaged and recorded as  $\bar{D}(F)$ . The equivalent force-extension curve  $F(\bar{D})$  can be obtained by the inversion of the relation  $\bar{D}(F)$ .

It was shown<sup>28</sup> that the difference between results of isometric and isotensional experiments is essential when the system is small (short molecules). For long and flexible molecules the fluctuations of variables are negligible and the difference between the two statistical ensembles vanishes.

Our RNA pseudoknot folding thermodynamics theory developed above can be further developed to study the thermodynamics of the mechanical unfolding of RNA pseudoknot. We assume that the two ends of the molecule are attached to two beads whose positions and the acting force are controlled by the force-extension measuring device. In our calculations, we consider the fixed force and fixed extension ensembles separately.

We assume that one end of the chain is fixed and the

constant force  $\mathbf{f}$  is applied to the other end in the constant force experiments. The average end-end vector  $\mathbf{D}$  of the molecule is recorded as a function of  $\mathbf{f}$ . The work  $\mathbf{f} \cdot \mathbf{D}$  done on the molecule by force  $\mathbf{f}$  contributes to the energy of the molecule, and the partition function for the constant force ensemble is

$$Z(T, \mathbf{f}) = \sum_E \sum_{\mathbf{D}} g(E, \mathbf{D}) e^{-(E - \mathbf{f} \cdot \mathbf{D})/k_B T}, \tag{6}$$

where  $g(E, \mathbf{D})$  is the constrained density of states, i.e., the number of conformations with the energy  $E$  and end-end vector  $\mathbf{D}$ . From  $Z(T, \mathbf{f})$ , the mean extension of the chain at the given force can be calculated as

$$\bar{\mathbf{D}}(T, \mathbf{f}) = k_B T \frac{\partial}{\partial \mathbf{f}} \ln Z(T, \mathbf{f}). \tag{7}$$

In the constant extension experiments, the end-end vector  $\mathbf{D}$  of the chain is assumed to be held constant and the

average force  $\mathbf{f}$  acting on the molecule is recorded as a function of  $\mathbf{D}$ . Since  $\mathbf{D}$  is constant, the partition function for the constant extension ensemble is

$$Z(T, \mathbf{D}) = \sum_E g(E, \mathbf{D}) e^{-E/k_B T}. \quad (8)$$

If  $Z(T, \mathbf{D})$  is known, the mean force  $\bar{\mathbf{f}}$  corresponding to the given  $\mathbf{D}$  can be calculated by the formula

$$\bar{\mathbf{f}}(T, \mathbf{D}) = -k_B T \frac{\partial}{\partial \mathbf{D}} \ln Z(T, \mathbf{D}). \quad (9)$$

## A. Density of states

In both the constant force and constant extension models, the key problem is how to calculate the density of states  $g(E, \mathbf{D})$ , which is the total number of conformations with energy  $E$  and end-end vector  $\mathbf{D}$ . We assume that one end of the molecule is attached to the wall and the force  $\mathbf{f}$  acting on the molecule is directed perpendicular to the wall. Instead of the end-end vector  $\mathbf{D}$ , we consider its component denoted as  $D$ , in the direction of the force. In the partition function calculation, the conformational ensemble includes both secondary structures and pseudoknotted structures. To find the number of conformations with a tail attached to the wall and the end of the other tail being at the distance  $D$  from the wall, we use the following approach. The structure can be divided into three parts: two tails and the structure element. For a given  $D$ , we exhaustively enumerate all the possibilities of three numbers:  $D_1$  (= extension of tail 1),  $D_p$  (= end-end extension of the structure element), and  $D_2$  (= extension of tail 2), such that  $D_1 + D_p + D_2 = D$ . For each set of  $(D_1, D_p, D_2)$ , we calculate the product of the numbers of conformations of the first tail with extension  $D_1$ , of the structure element with extension  $D_p$ , and of the second tail with extension  $D_2$ . In what follows we give detailed description for the calculation for pseudoknotted and secondary structures, separately.

### 1. Pseudoknotted conformations

To illustrate the principle, we choose the 2D lattice representation for chain conformations. On the two-dimensional lattice, a pseudoknot structure element (without tails) can have four different orientations with respect to the wall (Fig. 9). Since we consider the pseudoknot to be the central part of the structure, a convenient choice of the coordinate system  $(x, y)$  would be such that the  $y$  axis is along the stem with the tail attached to the wall. This is the same coordinate system as the one used for the calculation of the number of pseudoknot conformations (Fig. 1, Sec. III). In such coordinate system, the wall and the force have different orientations with respect to the pseudoknot (Fig. 9).

For a given orientation of the pseudoknot, we need two different functions for the conformational count of tails:  $\Omega_T^{\parallel}(t, d)$  and  $\Omega_T^{\perp}(t, d)$ , which denote the numbers of conformations for a tail of length  $t$  with end-end distance  $d$  in the direction parallel ( $\parallel$ ) and perpendicular ( $\perp$ ) to the first (closest to the pseudoknot structure element) bond of the tail,

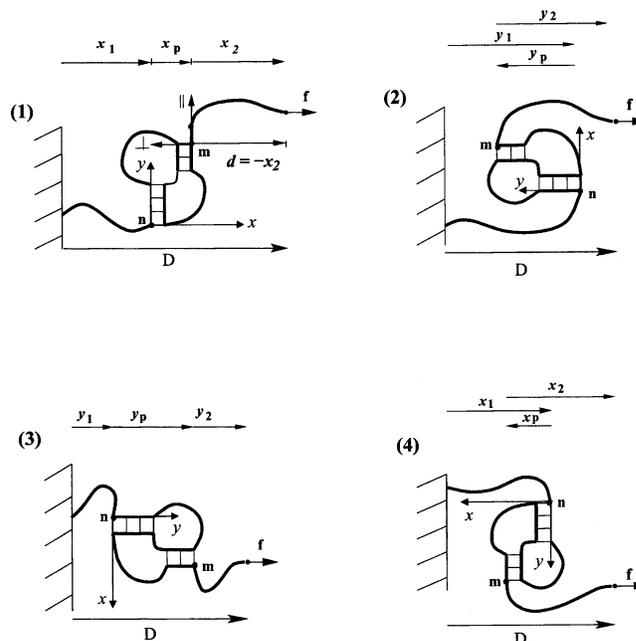


FIG. 9. In the two-dimensional lattice model, a pseudoknot can have four possible orientations with respect to the wall.  $x_p = x(m) - x(n)$  and  $y_p = y(m) - y(n)$ . The coordinate system can be defined by the position of the first stem of the pseudoknot ( $y$  axis along the stem), while the  $\parallel$  and  $\perp$  axes of the tail is defined by the orientation of the first bond as shown in (1).

respectively. We have obtained the  $\Omega_T^{\parallel}$  and  $\Omega_T^{\perp}$  functions by means of exact computer enumeration for  $1 \leq t \leq 11$ .

For a given extension  $D$  along the line of the action of force, we follow the following procedure to calculate the density of states (the number of conformations) for the pseudoknots.

1. We enumerate the different orientations of the pseudoknot, as shown in Fig. 9.
2. For each pseudoknot orientation, we enumerate the conformations of the pseudoknot enlarged interface. For a given enlarged interface conformation, the two helix stems, the energy  $E$  of the pseudoknot, and the end-end vector ( $x_p$  and  $y_p$  in Fig. 9) and the extension  $D_p$  of the pseudoknot element are all fixed.
3. For each enlarged interface conformation, using Eq. (3), we calculate the number of pseudoknot conformations  $\Omega_p$ .
4. We enumerate all the possibilities of  $D_1$  and  $D_2$  such that  $D_1 + D_2 = D - D_p$ . The numbers of tail 1 and tail 2 conformations which correspond to the given  $D_1$  and  $D_2$ , respectively, are approximated by the function  $\Omega_T^{\parallel}(t, d)$  or  $\Omega_T^{\perp}(t, d)$ . Which of two  $\Omega_T$  functions should be used depends on the orientations of the wall and the force with respect to the pseudoknot (cases 1–4 in Fig. 9) and on the directions of the first bonds of the tails. For example, for the first bond of tail 2 directed upward in Fig. 9 case 1, the number of tail 2 conformations which have the end-end extension  $D_2$  along the  $x$  axis, equals  $\Omega_T^{\perp}(t_2, d = -x_2)$ , because tail 2 has the end-end extension  $d = -D_2 = -x_2$  along the tail's (local)  $\perp$  axis (see Fig. 9).
5. For each pair of  $(D_1, D_2)$ , we calculate the product of

the number of conformations  $\Omega_P$  of the pseudoknot with the given enlarged interface and the numbers of conformations of the tails  $\Omega_T^z(t_1, d_1)$  and  $\Omega_T^z(t_2, d_2)$ , where  $t_1$  and  $t_2$  are the lengths of the two tails,  $z=\parallel$  or  $\perp$ , and  $d_1$  and  $d_2$  are required extensions of tails in direction  $z$ .

6. The density of states  $g_P(E, D)$  for pseudoknotted structures with energy  $E$  and end-end distance  $D$  is given by the sum over (i) all the possible  $(D_1, D_2)$ , (ii) the enlarged interface conformations  $\mathbf{I}$ , and (iii) the pseudoknot orientations,

$$g_P(E, D) = \sum_{\text{orientation}} \sum_{\mathbf{I}} \sum_{D_1} \sum_{D_2} \Omega_T(t_1, d_1) \cdot \Omega_P \cdot \Omega_T(t_2, d_2). \quad (10)$$

Here the value of parameters  $d_1$  and  $d_2$  are determined by the orientation of the pseudoknot with respect to the wall and by the enlarged interface conformation  $\mathbf{I}$ , which defines the direction of the first tail bonds with respect to the pseudoknot.

For the number of conformations  $\Omega_T$  for the single-stranded chain segments, we have exactly enumerated the conformations for different chain lengths  $10 \leq l \leq 28$  and different extensions  $D \leq l$ . We find that we can fit the results by two different functions:

$$\ln \Omega_T(l, D) = (-2.34782 + 0.803224l) + (-0.0239763 + 2.59832/l)D \quad (11)$$

for  $D \leq l/2$  (less stretched chain) and

$$\ln \Omega_T(l, D) = \left[ \frac{3.1649436}{l} - \frac{4.19464}{l^2} \right] D(l - D) \quad (12)$$

for  $l/2 \leq D \leq l$  (more stretched chain). In Fig. 10(c), we show the comparison between the above approximated  $\Omega_T(l, D)$  and that from exact computer enumeration and find good agreement. The above approximations for  $\ln \Omega_T(l, D)$  are used as extrapolations for longer open chains.

## 2. Secondary structures

We first use hairpin conformations to illustrate the methodology. We will then generalize the theory to treat more complex secondary structures. A unique feature of a hairpin structure element with tails is that the end-end distance (between monomers  $a$  and  $b$  in Fig. 10) is fixed. So the conformational count for a given total extension  $D$  can be calculated as

$$g_2(E, D) \approx k \Omega_H(E) \cdot \Omega_T(l, D), \quad (13)$$

where  $\Omega_H(E)$  is the number of conformations of the hairpin element (from  $a$  to  $b$  without tails) with energy  $E$ ,  $\Omega_T(l, D)$  is the number of conformations of the single-stranded chain segment as a function of length  $l = t_1 + t_2 + 1$  and the extension  $D$  [see Fig. 10(a)], and the prefactor  $k$  accounts for the volume exclusion between the tails and the hairpin structure element. We find that  $k \approx 1/4$  for  $D \leq l - 4$  and  $k \approx 1$  otherwise.

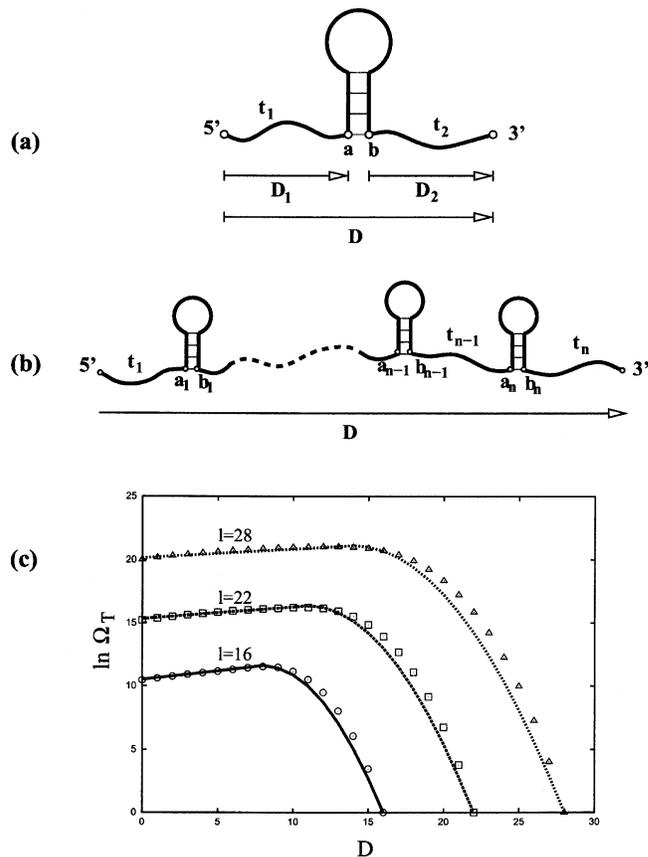


FIG. 10. (a) A hairpin (with tails) can be divided into hairpin element and the single-stranded segment, which consists of two tails ( $t_1$  and  $t_2$ ) and the outermost contact ( $a, b$ ) of the hairpin structure element. (b) A general secondary structure can be decomposed as multiple sequentially connected secondary structure elements. (c) The number of conformations of a single-stranded segment as a function of the segment's length  $l$  and the end-end extension  $D$  from exact enumeration (symbols) and the analytical approximation (lines).

For secondary structures, which consist of multiple sequentially connected secondary structure elements (e.g., hairpins), we calculate the density of states from the following recursive relation [see Fig. 10(b)]:

$$g_2^{(n)}(E, D) = k \Omega_H^{(n)}(E_n) g_2^{(n-1)}(E - E_n, D), \quad (14)$$

where  $E = \sum_{i=1}^n E_i$  is the total energy,  $E_i$  is the energy of the  $i$ th secondary structure element (between  $a_i$  and  $b_i$ ), and  $g_2^{(n-1)}$  is the density of states of a reduced chain, where the  $n$ th structure element (between  $a_n$  and  $b_n$ ) is replaced by a single bond connecting  $a_n$  and  $b_n$ .  $\Omega_H^{(n)}(E_n)$  is the number of conformations of the  $n$ th secondary structure element.

The sum of the densities of states for secondary and pseudoknot structures gives the (total) density of states  $g(E, \mathbf{D})$  in Eqs. (6) and (7). In Fig. 11, we show the tests for our theory against exact computer enumeration for a 38-mer pseudoknot-forming chain. We find good agreements.

## B. Force-extension curve, misfolded pseudoknots, and folding thermodynamics

We consider two specific pseudoknot-forming nucleotide sequences for which thermal unfolding processes have been studied (Figs. 7 and 8). In the calculation for the density of

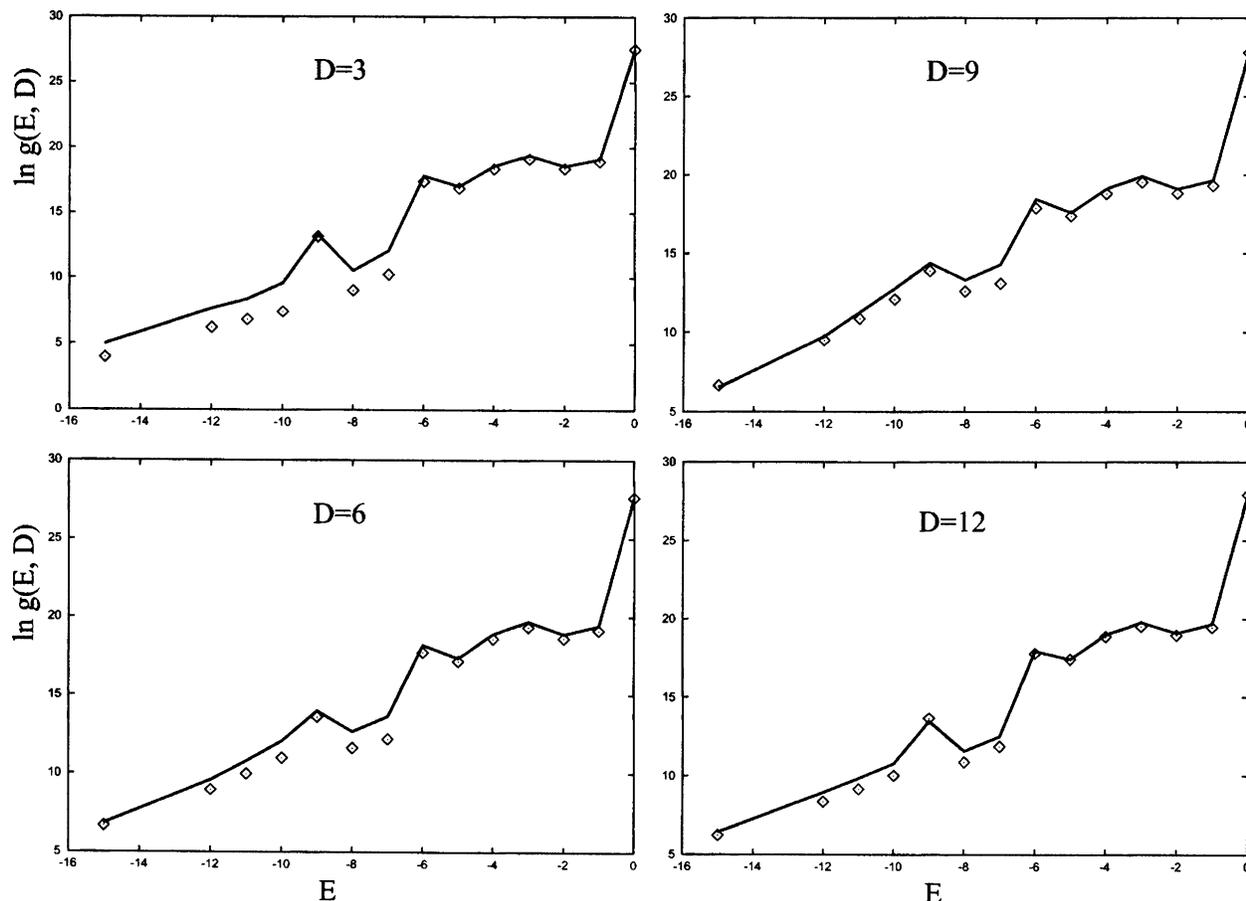


FIG. 11. The density of states  $g(E, D)$  for the pseudoknot/hairpin/open chain conformational ensemble for the 38-mer pseudoknot [shown in Fig. 4(a) with added tails of lengths  $t_1=3$  and  $t_2=4$ ]. Symbols: from exact enumeration; lines: from the theory developed in this study.

states, we enumerate all the possible secondary and pseudoknotted states, including all the possible misfolded states and partially folded states. We show the predicted force-extension curves and the conformational changes for the two sequences in Fig. 12.

We find that for the two sequences, both the isometric and isotensional curves show a major transition from the native pseudoknot ( $N$ ) to a misfolded intermediate state ( $I_1$ ). A notable feature shown in Fig. 12 is that the misfolded intermediate ( $I_1$ ) emerges when the extension of the molecule is both small and large compared to the average extension of the native pseudoknot  $N$ . The formation and the reformation of the intermediate state  $I_1$  (with both small and large extensions) can be explained in the following way.

The native pseudoknots have shorter tails than the intermediate states. Therefore, the end-to-end extension of the native pseudoknot is much more restricted than that of the intermediate state  $I_1$ . For instance, for sequence 1 in Fig. 12(a), the end-to-end extension for the native pseudoknot  $N$  and the intermediate  $I_1$  in a two-dimensional lattice are in the ranges of [3, 10] and [1, 13], respectively. Therefore, for sequence 1, for extension outside the range of [3,10], the native pseudoknot cannot exist and new structure would emerge. In contrast, the intermediate state  $I_1$  can accommodate a wide range of extension and can exist as a stable state outside the range of [3,10].

As shown in Fig. 10(c), the number of tail conformations

quickly decreases as the tail is stretched. When the native pseudoknot, which has short tails, is restricted to have very small or large extensions, the tails would inevitably be stretched. The entropy of the tails is small and the free energy  $F=E-TS$  of the structure is high. Therefore, the native pseudoknot is unstable and a structural transition would occur. For the misfolded structures ( $I_1$ ), however, the longer tails allow the chains to achieve the required (small or large) extensions without becoming highly stretched. So the intermediate state can be more stable than the native pseudoknots for small and large extensions.

In the limiting regime of very small end-to-end extension, the chain is in highly compact (and thus low-entropy) states. Large extension would cause an increase in the freedom of such highly compressed chain. So in the small extension limit, larger extension would lead to lower free energy, resulting in an apparent negative equilibrium pulling force.

Another notable feature in Fig. 12 is that sequence 1 mechanically unfolds through very different steps of conformational transitions than thermal unfolding. The thermal unfolding involves simple disruptions of native contacts, while in the mechanical unfolding, the native pseudoknot unfolds, refolds, and unfolds again during the pulling process. In contrast, for sequence 2, the intermediate state is the same misfolded hairpin as that in the thermal unfolding.

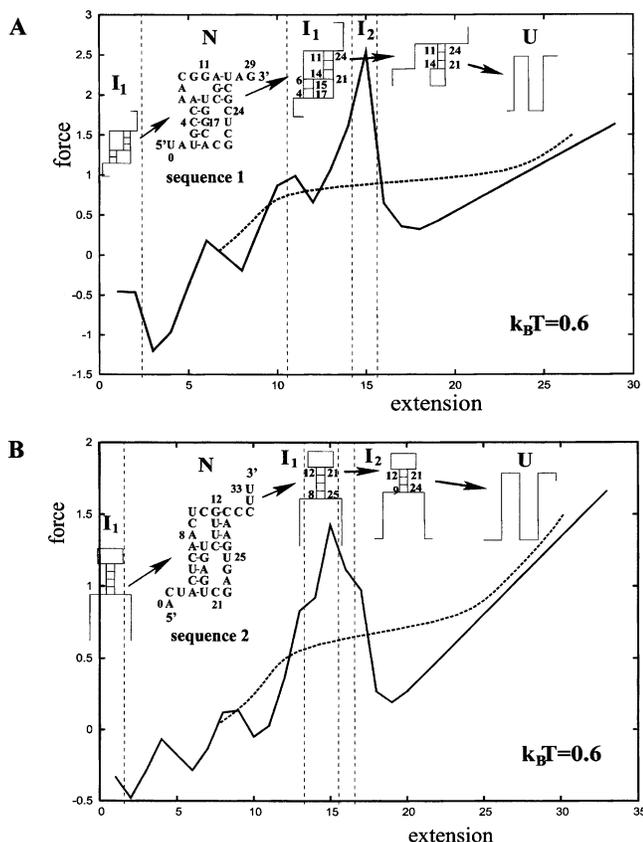


FIG. 12. The calculated isometric (solid lines) and the isotensional (dashed lines) force-extension curves and the structural transitions for the two nucleotide sequences specified. For both sequences, the misfolded structures ( $I_1$ ) emerges as intermediate states for those extensions of the molecules at which the overstretched native pseudoknots are less stable.

## VI. CONCLUSION AND DISCUSSION

We present a statistical mechanical theory for the folding thermodynamics of pseudoknotted chain conformations, including all the possible partially folded and misfolded structures. The model enables the calculation of conformational entropy and the partition function of pseudoknots. The key idea of the theory is (i) to decompose a pseudoknot structure into stem-loop subunits, (ii) to separate out the interfacial segments between the subunits, and (iii) to account for the correlation and volume exclusion between subunits through localized effects within and near the interface. The theory has been shown to give good predictions for the folding thermodynamics as tested against exact computer enumerations. The theory enables predictions for the folding stability, natively like and misfolded folding intermediates, and folding free energy landscapes for pseudoknotted and secondary structures.

The current form of the model is not without limitations. Possible further development of the model should address the following issues.

1. More realistic off-lattice chain representations can be used within the present graph-theoretic framework. An example of the possible off-lattice representation is the virtual bond model,<sup>29,30</sup> where the six bonds of each nucleotide backbone are replaced by two virtual bonds. The length of the virtual bond is nearly fixed (3.9 Å),<sup>30</sup>

and the conformation of each nucleotide is described by the torsional angles of the virtual bonds.

2. Noncanonical intraloop interactions and other tertiary interactions can play important roles in the sequence and temperature dependence of the loop entropy and the folding thermodynamics and therefore should be taken into account in the further development of the model. The generality of the basic ideas in the current model suggests the possibility to systematically extend the model to treat more complex tertiary folds, including ones with multiple crossing-linked contacts.<sup>8</sup>
3. The helical stems in pseudoknots tend to form the energetically favorable coaxial stacks. Such coaxial stacking is neglected in the present model and should be included in the further development of the model.
4. RNA folding is strongly dependent on the ionic solution condition. To properly take into account this effect, the chain conformational model should be combined with the polyelectrolyte theory to account for the ion electrostatics.

Thermodynamic parameters (especially the conformational entropy) for pseudoknotted structures are currently very limited, mainly due to the lack of a rigorous statistical mechanical model for pseudoknot folding. The present model provides a general method for the computation of the conformational entropy for pseudoknots. Moreover, the statistical mechanical framework developed here can also be used to extract the thermodynamic parameters from the experiments.

## ACKNOWLEDGMENT

This research was supported by the NIH through Grant No. GM063732 (to S.-J.C.).

- <sup>1</sup>D. H. Turner, N. Sugimoto, and S. M. Freier, *Annu. Rev. Biophys. Chem.* **17**, 167 (1988).
- <sup>2</sup>S. M. Freier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373 (1986).
- <sup>3</sup>J. A. Jaeger, D. H. Turner, and M. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7706 (1989).
- <sup>4</sup>A. P. Gulyaev, F. H. D. van Batenburg, and C. W. A. Pleij, *RNA* **5**, 609 (1999).
- <sup>5</sup>H. Jacobson and W. H. Stockmayer, *J. Chem. Phys.* **18**, 1600 (1950).
- <sup>6</sup>J. P. Abrahams, M. van den Berg, E. van Batenburg, and C. Pleij, *Nucleic Acids Res.* **18**, 3035 (1990).
- <sup>7</sup>A. Lucas and K. A. Dill, *J. Chem. Phys.* **119**, 2414 (2003).
- <sup>8</sup>Z. Kopeikin and S.-J. Chen, *J. Chem. Phys.* **122**, 094909 (2005).
- <sup>9</sup>K. M. Fiebig and K. A. Dill, *J. Chem. Phys.* **98**, 3475 (1993).
- <sup>10</sup>S.-J. Chen and K. A. Dill, *J. Chem. Phys.* **103**, 5802 (1995).
- <sup>11</sup>S.-J. Chen and K. A. Dill, *J. Chem. Phys.* **114**, 4602 (1998).
- <sup>12</sup>W. Zhang and S.-J. Chen, *J. Chem. Phys.* **114**, 7669 (2001).
- <sup>13</sup>I. Tinoco, Jr., *Annu. Rev. Biophys. Biomol. Struct.* **33**, 363 (2004).
- <sup>14</sup>X. Zhuang and M. Rief, *Curr. Opin. Struct. Biol.* **13**, 88 (2003).
- <sup>15</sup>C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, *Curr. Opin. Struct. Biol.* **10**, 279 (2000).
- <sup>16</sup>M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez, *Prog. Biophys. Mol. Biol.* **74**, 63 (2000).
- <sup>17</sup>M. S. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, *Science* **276**, 1112 (1997).
- <sup>18</sup>L. Tskhovrebova, J. Trinick, J. A. Sleep, and R. M. Simmons, *Nature (London)* **387**, 308 (1997).
- <sup>19</sup>U. Bockelmann, P. Thomen, B. Essevez-Roulet, V. Viasnoff, and F. Heslot, *Biophys. J.* **82**, 1537 (2002).
- <sup>20</sup>J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **292**, 733 (2001).

- <sup>21</sup>J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **296**, 1832 (2002).
- <sup>22</sup>B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **299**, 1892 (2002).
- <sup>23</sup>S. Harlepp, T. Marchal, J. Robert, J. F. Leger, A. Xayaphoumimine, H. Isambert, and D. Chatenay, *Eur. Phys. J. E* **12**, 605 (2003).
- <sup>24</sup>S. Cocco, R. Monasson, and J. F. Marko, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8608 (2001).
- <sup>25</sup>A. Montanari and M. Mezard, *Phys. Rev. Lett.* **86**, 2178 (2001).
- <sup>26</sup>U. Gerland, R. Bundschuh, and T. Hwa, *Biophys. J.* **81**, 1324 (2001).
- <sup>27</sup>U. Gerland, R. Bundschuh, and T. Hwa, *Biophys. J.* **84**, 2831 (2003).
- <sup>28</sup>D. Keller, D. Swigon, and C. Bustamante, *Biophys. J.* **84**, 733 (2003).
- <sup>29</sup>W. K. Olson, *Macromolecules* **13**, 721 (1980).
- <sup>30</sup>S. Cao and S.-J. Chen, *RNA* **11**, 1884 (2005).