Statistical thermodynamics for chain molecules with simple RNA tertiary contacts

Zoia Kopeikin and Shi-Jie Chen[a)]
Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211
and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211

(Received 8 December 2004; accepted 16 December 2004; published online 28 February 2005)

A statistical thermodynamic model is developed for chain molecules with simple RNA tertiary contacts. The model, which accounts for the excluded volume effect and the nonadditivity in the free energy, enables reliable predictions for the conformational entropy and partition function for simple tertiary folds. Illustrative applications are made to conformational transitions involving simple tertiary contacts. The model can predict the interplay between the secondary and the tertiary interactions in the conformational changes. Though the present form of the theory is tested and validated in a two-dimensional lattice model, the methodology, which is developed based on a general graphical representation for chain conformations, is applicable to any off-lattice chain representations. Moreover, the analytical formulation of the method makes possible the systematic development of the theory for more complex tertiary structures. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1857831]

I. INTRODUCTION

Biopolymers fold into compact three-dimensional structures. For ribonucleic acid molecules (RNAs), we can distinguish the structures at the level of secondary and tertiary structures. Secondary structures consist of a list of intrachain contacts (=base pairs) such that any two contacts (i,j) $(i<j)$ and (k,l) $(k<l)$ are either nested $(i<k<l<j)$ or unrelated $(i<j<k<l)$. According to the intrachain contacts, we can decompose a secondary structure into relatively independent secondary structural motifs (subunits), such as helices
(=stacked base pairs) and loops. Tertiary structures consist of intrachain contacts (i,j) $(i<j)$ and (k,l) $(k<l)$ that are crossing-linked $(i<k<j<l)$. The tertiary structure describes how the different parts of the secondary structures are brought together by the tertiary crossing links to form a compact three-dimensional structure. According to such definitions for the secondary and tertiary structures, probably the simplest tertiary structure is an RNA pseudoknot, which contains nucleotides in the loop paired with nucleotides external to the loop.

Different subunits of a RNA secondary structure are not crossing linked, so they are relatively independent of each other. As a result, the total free energy can be approximately computed as an additive sum of the free energy of each subunit.[1–3] Such additivity makes it possible to use the computationally efficient dynamic programming to predict the native structure and the equilibrium thermodynamics for secondary structures.[4,5] Based on the free energy additivity and the experimentally measured parameters for individual base stacks and loops,[6–8] several statistical mechanical models have been developed for RNA secondary structures.[5,9–17] A recent model goes beyond the additivity assumption by considering the excluded volume interferences between neighboring subunits.[12–16] The nonadditive model gives the partition function as a product of the matrix-form partition function of each component subunit. The model gives good predictions for the melting curves and the free energy landscapes.[14]

Though RNA tertiary interactions, which are formed in the late stage of folding, are generally weak as compared to interactions in secondary structures, they play a variety of crucial structural and functional roles. Despite the extensive experimental[18–28] and computational studies[29–32] for the pseudoknots and tertiary structures, our ability to model RNA tertiary folding thermodynamics is still very limited. This is mainly due to two interrelated problems: the lack of energy parameters and the lack of a statistical mechanical model. Experimentally, it is hard to obtain the tertiary energy parameters because the crossing links in the tertiary structure cause strong correlations between different structural subunits, which result in a very convoluted interplay between the secondary and tertiary interactions and between the different tertiary interactions. We need a model to deconvolute the experimental results and to extract the thermodynamic parameters from the experimental measurements.

Due to the interdependence of different structural subunits, the conformational entropy and the free energy become nonadditive, i.e., the conformational entropy and the free energy of the chain are not equal to the sum of the conformational entropy and the free energy of each individual subunit. As a result, the parameters for individual loops and stacks alone are not sufficient for the prediction of pseudoknot thermodynamics. We need a model that can account for the coupling between different structural subunits. In this paper, we present a statistical mechanical model that can treat such coupling through simple tertiary contacts. The

[a)]Author to whom correspondence should be addressed. Electronic mail: ChenShi@missouri.edu

development of such a model would enable us to extract the thermodynamic parameters from the thermodynamic experiments.[33,34]

A major challenge in the prediction of the tertiary and pseudoknot thermodynamics is how to evaluate the loop entropy. Several attempts have been made to compute the loop entropy for RNA pseudoknots.[29,31,32,35,36] For example, a set of thermodynamic parameters for RNA *H*-pseudoknots has been proposed by Gultyaev, Batenburg, and Pleij.[31] Used in structure prediction programs, the parameters can give a better approximation then the previous one.[30] These models used the Stockmayer–Jacobson approximation, where the excluded volume (self-avoiding) effect and the nonadditivity effect are not explicitly considered. A recent statistical mechanical model[36] can give accurate prediction for the thermodynamics of the canonical pseudoknots with two stems and two loops. In the present study, we focus on general conformations with simple tertiary contacts, for example, the one where two secondary structures are brought into contact by simple tertiary contacts, so the conformations addressed in the model include a broader class of tertiary folds. Moreover, the present theory is developed according to the increasing conformational complexity, therefore, it has the potential to be systematically generalized to treat more complex tertiary folds.

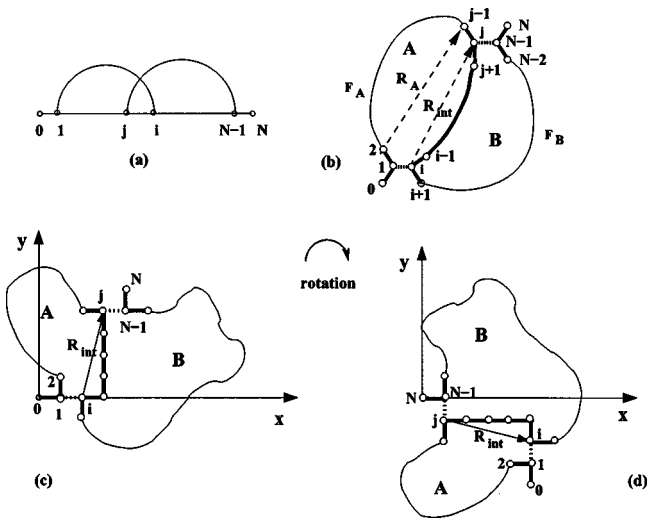

FIG. 1. (a) The simplest graph with two crossing links. (b) The crossing-linked chain conformation consists of two loops *A* and *B* having the common chain segment (interface). It can be divided into relatively independent subunits: enlarged interface (shown bold) and two free single-stranded segments (F_A and F_B). The number of conformations of segment F_A can be considered to be a function of its end-end vector R_A, or, as a further approximation, of interfacial end-end vector R_{int}. (c) and (d) Two orientations of the enlarged interface on two-dimensional lattice with the given coordinate axes. The coordinate systems in (c) and (d) define the arguments of the function $\omega_f(y_{\text{int}}, l_{\text{int}}, f)$ for the numbers of conformations of segment F_A and F_B, respectively.

II. POLYMER GRAPH AND PARTITION FUNCTION

Our approach is based on the graphic representation of intrachain contacts.[12,13,16,37] A polymer graph consists of vertices for the monomers and straight line links for the covalent bonds connecting the monomers. Monomers in spatial contact (base pairs in RNA) are also connected by curved links. A polymer graph represents an ensemble of chain conformations that contain the intrachain contacts specified by the curved links in the graph. There are three possible relationships between two contacts (curved links): nested, unrelated, and crossing linked. Secondary structures involve only nested and unrelated contacts. In the present study, we focus on conformations having tertiary interactions, i.e., having crossing links in the polymer graph.

In terms of the polymer graph, the partition function can be calculated as a sum over all possible graphs,

$$Q(T) = \sum_{\text{graph}} \Omega e^{-E/k_B T}, \tag{1}$$

where k_B is Boltzmann's constant, T is the temperature, E is the energy, and Ω is the number of viable chain conformations for the graph. We assume that the intrachain contacts determine the energy E, as a result, a graph represents an equal-energy conformational ensemble. A central problem in the calculation of the partition function and the prediction of the statistical thermodynamics is the evaluation of the conformational count Ω.

In the present study, we focus on a large class of graphs (conformations) that involve simple tertiary contacts, namely, graphs formed by adding simple curved (tertiary) crossing link(s) to an otherwise secondary structure polymer graph. In the model development, in order to illustrate the theory, we will systematically add tertiary links so that they cross one, two, three, and more links in the otherwise noncrossing links. We compute the conformational entropy for each graph and the partition function $Q(T)$ for all the possible secondary structures (noncrossing contacts) and all the tertiary structures that can be represented by simple crossing links. Our theory is based on a graphical representation for the intrachain contacts (base pairs), so the general methodology does not depend on any specific chain representation. However, for the purpose of illustration, we use two-dimensional lattice to represent chain conformations.

III. CONFORMATIONS WITH TWO CROSSING LINKS

We first treat conformations that can be represented by graphs that contain only two crossing-linked contacts. Such type of graphs represent a large class of conformations because the graphs can contain an arbitrary number of nested or unrelated contacts and so in general can form complex secondary structures. In the conformational space, the crossing links provide linkers for the different secondary structures.

A. Basic graphs with two crossing links

We start with the simplest elementary graphs that contain only two contacts, which are crossing linked; see contacts $(1,i)$ and $(j,N-1)$ in Fig. 1(a). Our goal in this section is to develop a theory to count chain conformations for such a graph. To focus on how the curved links affect the conformational statistics, we neglect the dangling tails, which involve no curved links. We use only single monomers [labeled as 0 and N in Figs. 1(a) and 1(b)] to account for the excluded volume interactions between the tails and the

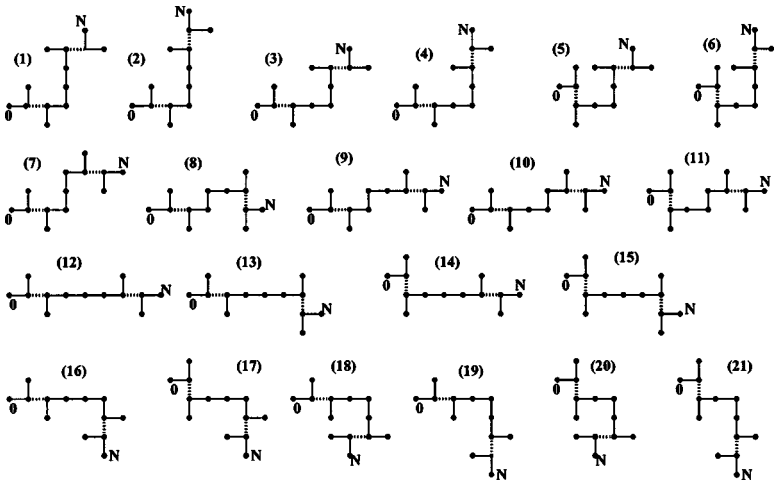

FIG. 2. The viable conformations of the 5-mer interface, numbered from 1 to 21.

linked part of the conformation. The full tails will be added back in the final partition function calculation to keep the completeness of the conformational ensemble.

As illustrated in Fig. 1(b), the whole chain conformation can be decomposed into two loops *A* and *B*. Loops *A* and *B* are correlated through (1) the interface from monomer j to monomer i, namely, the (common) chain segment $j \rightarrow i$ should have exactly the same conformation in both loops, and (2) the steric hindrance (the excluded volume effect) between the two loops. As a result, the total number of the chain conformations is usually much smaller then the product of the number of conformations of each individual (isolated) loop.

In general, exact calculation for the excluded volume interactions between *A* and *B* is not viable. However, since the excluded volume interferences between *A* and *B* occur mainly near the interface, we can focus on the interface, which is a much simpler and a much more manageable system. To account for the excluded volume in the vicinity of the interface, we define an "enlarged interface" as the system consisting of the interface and its neighboring monomers. Specifically, as shown in Fig. 1(b), the enlarged interface **I** consists of monomers 0, 1, 2 and $N-2$, $N-1$, N in addition to the chain segment from monomer $j-1$ to monomer $i+1$. The enlarged interface can approximately account for the correlations between *A* and *B*. To describe the conformations for the other parts of the chain, we define "free loop segment" F_A for chain segment from monomer 2 to $j-1$ for loop *A* and free loop segment F_B for chain segment from monomer $i+1$ to $N-2$ for loop *B*. F_A and F_B have chain lengths of $f_A=j-3$ and $f_B=N-i-3$, respectively. For each given conformation **I** of the enlarged interface, we use $\Omega_f(f_A,\mathbf{I})$ and $\Omega_f(f_B,\mathbf{I})$ to denote the number of conformations for F_A and F_B for a given **I**.

With the separation of the free loop segments F_A and F_B from the enlarged interface **I**, the computation of the number of accessible conformations Ω for the two-contact graph in Fig. 1(a) becomes tractable, and Ω can be computed as

$$\Omega = \sum_{I=1}^{\omega_I} \Omega_f(f_A,\mathbf{I}) \cdot \Omega_f(f_B,\mathbf{I}), \tag{2}$$

where $I=1,2,\ldots,\omega_I$ denotes the viable conformations of the enlarged interface, ω_I is the number of the viable conformations of the enlarged interface. Figure 2 shows all the $\omega_I=21$ conformations for an enlarged interface with a 5-mer (four-bonds) interface in a two-dimensional lattice. The number of the enlarged interface conformations ω_I increases exponentially with the chain length of the interface.

Central to the computation of Ω from Eq. (2) is the calculation of $\Omega_f(f_A,\mathbf{I})$ and $\Omega_f(f_B,\mathbf{I})$. In the following, in order to be specific, we use $\Omega_f(f_A,\mathbf{I})$ to illustrate the methodology. Due to the large number of possibilities for $(f_A,\mathbf{I})$, it is impractical to have a precomputed table that lists all the values of $\Omega_f(f_A,\mathbf{I})$ for all the possible $(f_A,\mathbf{I})$'s. However, as we will show in the following, through successive approximations, we can transform the calculation for $\Omega_f(f_A,\mathbf{I})$ into a much simpler and computationally viable problem.

1. General methodology

First, as shown in Fig. 1(b) for the free loop segment F_A and the enlarged interface **I**, $\Omega_f(f_A,\mathbf{I})$ mainly depends on the enlarged interface conformation **I** through the positions of the two ends of the enlarged interface (i.e., monomers 2 and $j-1$) and the excluded volume interactions between F_A and **I** in loop *A*. Therefore, for $\Omega_f(f_A,\mathbf{I})$ we can approximately represent the conformation **I** of the enlarged interface by the chain length of the interface $l_{\text{int}}(=i-j)$ and the end-end vector $\mathbf{R}_A$=the vector from monomer 2 to monomer $(j-1)$ [see Fig. 1(b)]. We note that $\mathbf{R}_A$ is also equal to the end-end vector of F_A. A larger end-end distance corresponds to more stretched conformations of **I** and F_A, and gives less accessible conformations, i.e., a smaller $\Omega_f(f_A,\mathbf{I})$. Through this approximation, we can compute $\Omega_f(f_A,\mathbf{I})$ as a function of $(\mathbf{R}_A,l_{\text{int}},f_A)$.

Second, for more complex graphs with multiple crossing links and multiple interfaces, it is hard to track the conformations for each of the enlarged interfaces. So it would be much more convenient to use the interfaces rather than the enlarged interfaces. Therefore, we simplify the $\mathbf{R}_A$ dependence by the $\mathbf{R}_{\text{int}}$ dependence, where $\mathbf{R}_{\text{int}}$ is the end-end vector of the interface [see Fig. 1(b)] instead of the enlarged interface. To account for the excluded volume effect, which is originally represented by the enlarged interface, we approximate the number of the conformations of the free loop segment $\Omega_f(f_A,\mathbf{I})$ for a given enlarged interface conformation **I** by the average $\overline{\Omega_f(f_A,\mathbf{I})}$ over all the possible enlarged

interface conformations that have the same end-end vector of the interface $\mathbf{R}_{\text{int}}$. The resultant approximate value of $\Omega_f(f_A,\mathbf{I})$ would be a function of $\mathbf{R}_{\text{int}}$ rather than of the enlarged interface conformation $\mathbf{I}$.

Third, because no intrachain contact exists for the interface chain segment, its conformation is largely extended. As an approximation, we assume that the interfacial chain does not double back. As a result, for $\mathbf{R}_{\text{int}}=(x_{\text{int}},y_{\text{int}})$, where $x_{\text{int}}=x_j-x_i$, $y_{\text{int}}=y_j-y_i$ in Fig. 1(b), we have

$$l_{\text{int}}=|x_{\text{int}}|+|y_{\text{int}}|. \tag{3}$$

The above relation shows that the x and y components of $\mathbf{R}_{\text{int}}$ are not independent of each other for a given l_{int}. Therefore, we can further simplify the dependence of $\Omega_f(f_A,\mathbf{I})$ on the *vector* $\mathbf{R}_{\text{int}}$ as the dependence on a single *scalar* variable, say, y_{int}. The y_{int} dependence of $\Omega_f(f_A,\mathbf{I})$ is much more manageable than the original $\mathbf{I}$ dependence because of the much less possibilities for y_{int} than for $\mathbf{I}$.

For a given vector $\mathbf{R}_{\text{int}}$ the actual values of the x_{int} and y_{int} components depend on the coordinate system. The coordinate system should be defined in such a way that it gives consistent treatment for loops A and B. The conformation of the enlarged interface defines the directionality of the coordinate system. For $\Omega_f(f_A,\mathbf{I})$, we define the coordinate system by fixing the coordinates of the first three monomers, labeled as 0, 1, and 2 in Fig. 1(c), to (0,0), (1,0), and (1,1). While for $\Omega_f(f_B,\mathbf{I})$, we use the equivalent monomers, namely, monomers N, $N-1$, and $N-2$, to define the coordinate system. The correspondence between these two sets of monomers becomes obvious if we apply a rotational transformation to the two loops, as shown in Fig. 1(d). For the enlarge interface conformation shown in Figs. 1(c) and 1(d), $\mathbf{R}_{\text{int}}=(x_{\text{int}},y_{\text{int}})$ is equal to (1, 4) for loop A [see Fig. 1(c)] and $(4,-1)$ for loop B [see Fig. 1(d)].

Finally, by averaging over all the possible interface chain conformations that have the same y_{int}, we can represent $\Omega_f(f_A,\mathbf{I})$ as a function of the y_{int} of the interface chain conformation instead of $\mathbf{R}_{\text{int}}$. The resultant $\Omega_f(f_A,\mathbf{I})$ for a given enlarged interface conformation $\mathbf{I}$ is simplified as a function of y_{int} (= the y component of the end-end vector of the interface $\cong$ that of the free loop chain segment), l_{int} (= the chain length of the interface), and f_A (= the length of the free loop chain segment):

$$\Omega_f(f_A,\mathbf{I})\simeq\omega_f(y_{\text{int}},l_{\text{int}},f_A). \tag{4}$$

For example, for the enlarged interface conformation in Figs. 1(c) and 1(d), l_{int} is equal to 5, and (y_{int},f_A) is equal to $(4,j-3)$ for loop A and $(-1,N-i-4)$ for loop B. Therefore, $\Omega_f(f_A,\mathbf{I})\simeq\omega_f(4,5,j-3)$; $\Omega_f(f_B,\mathbf{I})\simeq\omega_f(-1,5,N-i-4)$.

With Eq. (4) we have

$$\Omega\simeq\sum_I\omega_f(y_{\text{int}}^{(A)},l_{\text{int}}^{(A)},f_A)\omega_f(y_{\text{int}}^{(B)},l_{\text{int}}^{(B)},f_B), \tag{5}$$

where $(y_{\text{int}}^{(A)},l_{\text{int}}^{(A)},f_A)$ and $(y_{\text{int}}^{(B)},l_{\text{int}}^{(B)},f_B)$ are the respective parameter sets for a given conformation $\mathbf{I}$ of the enlarged interface.

In the limit of very short interface chain segment, i.e., if the interface is much shorter than the free loop chain segment, $l_{\text{int}}\ll f$, where f is the chain length of the free loop segment, $\Omega_f(f,\mathbf{I})$ would be only weakly dependent on the enlarged interface conformation $\mathbf{I}$. As a result, we can neglect the y_{int} dependence of $\omega_f(y_{\text{int}},l_{\text{int}},f)$ and approximate it by y_{int} independent function $\omega_0(l_{\text{int}},f)$:

$$\omega_f(y_{\text{int}},l_{\text{int}},f)\simeq\omega_0(l_{\text{int}},f)=\frac{\Omega_{\text{loop}}}{\omega_l}, \tag{6}$$

where Ω_{loop} is the number of conformations of the loop and ω_l is the number of conformations for the part of the enlarged interface within the loop. For example, in Fig. 1, for the free loop chain segment F_A from monomer 2 to monomer $j-1$, $f_A=j-3$ and $l_{\text{int}}=i-j$. Ω_{loop} is for the loop $1\to i\to 1$ and ω_l is for the chain segment $j-1\to i\to 1\to 2$, which is part of the enlarged interface. In terms of the $\omega_0(l_{\text{int}},f)$ function, for $l_{\text{int}}\ll f_A$, f_B, we can compute Ω in Eq. (5) as

$$\Omega\simeq\omega_l\omega_0(l_{\text{int}},f_A)\omega_0(l_{\text{int}},f_B). \tag{7}$$

2. Illustrative calculations and tests in two-dimensional lattice model

We choose a short (25-mer) chain to illustrate the method. The chain makes two crossing linked contacts, as specified by the graph in Fig. 3(a). For the given graph, the length of the interface $8\to 12$ is $l_{\text{int}}=4$, and the lengths of the free loop segments are $f=5$ and 9 for $2\to 7$ and $13\to 22$, respectively.

To compute the number of conformations Ω from Eq. (5), for each of the 21 enlarged interface conformations of the 5-mer interface in Fig. 2, we need to know $\omega_f(y_{\text{int}},l_{\text{int}},f)=\omega_f(y_{\text{int}},4,5)$ and $\omega_f(y_{\text{int}},4,9)$ for the respective free loop segments. Following the step-by-step procedure presented in the preceding section for the calculation of $\omega_f(y_{\text{int}},l_{\text{int}},f)$ function, we performed exact computer enumeration in a two-dimensional lattice and obtained $\omega_f(y_{\text{int}},l_{\text{int}},f)$ for all the possible parameter sets for $1\leqslant l_{\text{int}}\leqslant 12$ and $1\leqslant f\leqslant 24$. Table I shows the results for small l_{int} and f values.

For each conformation $\mathbf{I}$ of the enlarged interface in Fig. 2, we determine the y_{int} value of the interface in the coordinate system defined by the positions of monomers 0, 1, 2 for $y_{\text{int}}^{(A)}$ and by the positions of monomers N, $N-1$, $N-2$ for $y_{\text{int}}^{(B)}$. Summing over all the 21 conformations gives the number of chain conformations for the graph

$$\Omega=\sum_I\omega_f(y_{\text{int}}^{(A)},4,5)\omega_f(y_{\text{int}}^{(B)},4,9)=45.12. \tag{8}$$

The details of calculations are shown in Table II. The exact value for Ω obtained from the exact computer enumeration is 49, which is quite close to the estimated result.

In Fig. 3(c), as a test for the method, we compute the number of conformations for a series of graphs with two crossing links. We again find good agreement between the analytical calculation and the exact computer enumeration.

B. More complex graphs with two crossing links

The above theory can be generalized to treat more complex graphs that contain multiple noncrossing links in addi-

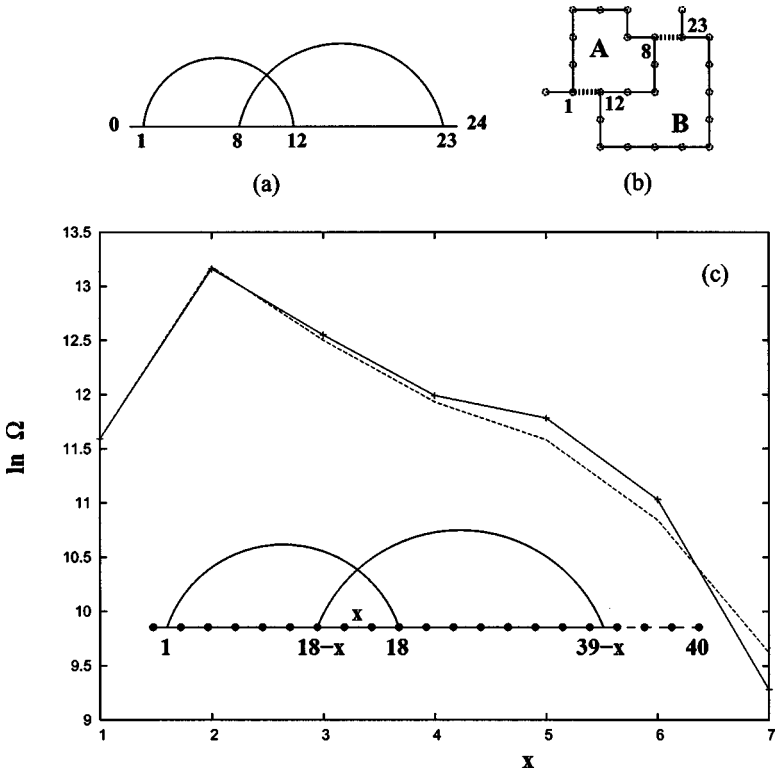

FIG. 3. (a) and (b) A simple graph with two crossing links is used to illustrate the method of calculation of values of function $\omega_f(y_{\text{int}}, l_{\text{int}}, f)$ (see Table I) and how to obtain the total number of conformations for the given graph (Table II). (c) To test our method, the set of graphs with variable position of the second contact has been chosen and the number of corresponding conformations as a function of x plotted. The approximation (dashed line) has been found to be in a good agreement with exact enumeration (solid line).

tion to two crossing links; see Fig. 4(a) for an example. The noncrossing links bear either nested or unrelated relationships with the crossing links. Since a cluster of the nested and unrelated links forms a secondary structure, the type of graphs in Fig. 4(a) can be regarded as (tertiary) crossing-linked secondary structures.

We use the graph in Fig. 4(b) to illustrate the theory. The difference between the graph/conformation in Figs. 4(b) and 1 comes from the additional loop A_1 attached to loop A_0 in Fig. 4(b). We treat the composite loop A_0+A_1 as an effective loop A. The effective "free loop segment" F_A for "loop A" is the chain segment from monomer 2 to monomer $j-1$, and the free loop segment F_B for loop B is from monomer $i+1$ to monomer $N-2$. The interface is from monomer j to monomer i of length $l_{\text{int}}=i-j$, and the enlarged interface is shown as the thick lines in Fig. 4(b).

We again use Eqs. (2) and (5) to treat the graph. The key is how to compute $\Omega_f(f_A, \mathbf{I})$ (= the number of conformations for F_A for a given conformation $\mathbf{I}$ of the enlarged interface). To account for the conformational constraint imposed by the additional pair (k,m) and the excluded volume interactions between A_0 and A_1, we classify four types of conformations for the contact (k,m) in a two-dimensional lattice[12] [see Fig. 4(c)]. For a given μth ($\mu=1,2,3,4$) type conformation of the (k,m) contact, we use $S_{A_1}^{(\mu)}$ to denote the number of conformations for loop A_1, and use $\omega_f^{(\mu)}(y_{\text{int}}, l_{\text{int}}, f_{A_0}) = \omega_f^{(\mu)}(y_j$

TABLE I. Values of $\omega_f(y_{\text{int}}, l_{\text{int}}, f)$ are shown for short interfaces and free loop segments. y_{int} varies from $(l_{\text{int}}-1)$ to $-(l_{\text{int}}-2)$. Free segments with even (odd) f can close only interfaces with odd (even) number of bonds.

l_{int}	y_{int}	$f=1$	2	3	4	5	6	7	8	9	10
1	0		0.00		1.00		1.00		1.00		5.00
2	1	0.00		0.00		0.00		0.00		0.00	
	0	0.00		0.50		1.00		2.50		6.25	
3	2		1.00		0.00		0.00		0.00		0.00
	1		0.00		0.00		2.00		4.00		9.00
	0		0.00		0.50		2.00		6.25		17.75
	−1		0.00		0.00		0.00		1.00		7.50
4	3	0.00		1.00		0.00		0.50		1.00	
	2	0.40		0.40		0.20		0.60		1.60	
	1	0.00		0.00		0.50		2.75		8.50	
	0	0.00		0.00		0.50		3.50		13.50	
	−1	0.00		0.00		0.00		0.00		1.00	
	−2	0.00		0.00		0.00		0.00		1.00	

TABLE II. The illustrative calculations of the number of conformations of the chain with two crossing links shown in Figs. 3(a) and 3(b). The pretabulated values of $\omega_f(y_{\rm int}, l_{\rm int}, f)$ have been used. The resulting value (45.12) is found to be in a good agreement with the exact conformational count $\Omega=49$.

I	$y_{\rm int}^{(A)}$	$y_{\rm int}^{(B)}$	$\omega_f(y_{\rm int}^{(A)},4,5)$	$\omega_f(y_{\rm int}^{(B)},4,9)$	Product
1	3	−1	0.0	1.0	0.0
2	3	−1	0.0	1.0	0.0
3	2	−2	0.2	1.0	0.2
4	2	−2	0.2	1.0	0.2
5	2	−2	0.2	1.0	0.2
6	2	−2	0.2	1.0	0.2
7	2	2	0.2	1.6	0.32
8	1	1	0.5	8.5	4.25
9	1	1	0.5	8.5	4.25
10	1	1	0.5	8.5	4.25
11	1	1	0.5	8.5	4.25
12	0	0	0.5	13.5	6.75
13	0	0	0.5	13.5	6.75
14	0	0	0.5	13.5	6.75
15	0	0	0.5	13.5	6.75
16	−1	3	0.0	1.0	0.0
17	−1	3	0.0	1.0	0.0
18	−2	2	0.0	1.6	0.0
19	−2	2	0.0	1.6	0.0
20	−2	2	0.0	1.6	0.0
21	−2	2	0.0	1.6	0.0
					Sum=45.12

$-y_i, i-j, k+j-m-2)$ to denote the number of conformations for the free loop segment F_{A_0} which consists of the chain segment $2\rightarrow k$, the contact (k,m), and the chain segment $m \rightarrow j-1$. F_{A_0} has chain length of $f_{A_0}=k+j-m-2$.

The sum over the four types of the (k,m) contact conformations gives $\Omega_f(f_A,\mathbf{I})$:

$$\Omega_f(f_A,\mathbf{I}) = \sum_{\mu=1}^{4}\sum_{\nu=1}^{4} \omega_f^{(\mu)}(y_j - y_i, i-j, k+j-m-2) Y_{\mu\nu} S_{A_1}^{(\nu)}, \tag{9}$$

where $Y_{\mu\nu}=1$ and 0 for a viable and nonviable connection between a type μ and a type ν conformation, respectively.[12] For example, $Y_{12}=1$, $Y_{24}=0$. Furthermore, through exact computer enumeration for $\omega_f^{(\mu)}$'s, we find that in a two-dimensional lattice

$$\alpha^{(\mu)} = \frac{\omega_f^{(\mu)}(y_{\rm int}, l_{\rm int}, f_{A_0})}{\omega_f(y_{\rm int}, l_{\rm int}, f_{A_0})} \simeq 0, 0.15, 0.15, 0.18$$

$$\text{for } \mu = 1,2,3,4, \text{ respectively.}$$

Here we assume that the total length of loop A_0 is longer than 4.

Combining the above results, we obtain the following simplified expression for $\Omega_f(f_A,\mathbf{I})$:

$$\Omega_f(f_A,\mathbf{I}) = \omega_f(y_{\rm int}, l_{\rm int}, f_{A_0}) \sum_{\mu=1}^{4}\sum_{\nu=1}^{4} \alpha^{(\mu)} Y_{\mu\nu} S_{A_1}^{(\nu)}. \tag{10}$$

For F_B, from Eq. (4), we have $\Omega_f(f_B,\mathbf{I})=\omega_f(y_{\rm int}, l_{\rm int}, f_B)$. With the above results for $\Omega_f(f_A,\mathbf{I})$ and $\Omega_f(f_B,\mathbf{I})$ we obtain the conformational count Ω from Eq. (2) for the graph.

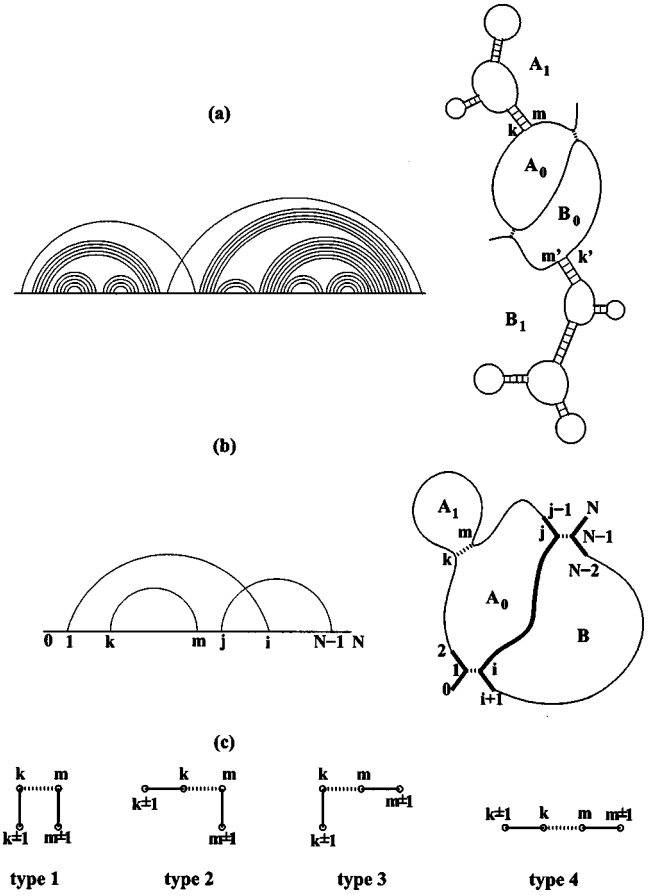

FIG. 4. (a) Two crossing-linked secondary structures. The number of conformations of each secondary structure can be computed from the previously developed matrix method (Refs. 12 and 13) to obtain $S_{A_1}^{(\nu)}$ in Eq. (9). (b) To demonstrate the method, the simplest representative of the secondary structure, loop A_1, has been attached to the loop A_0. (c) The four types of two-dimensional lattice conformations for a contact (k,m).

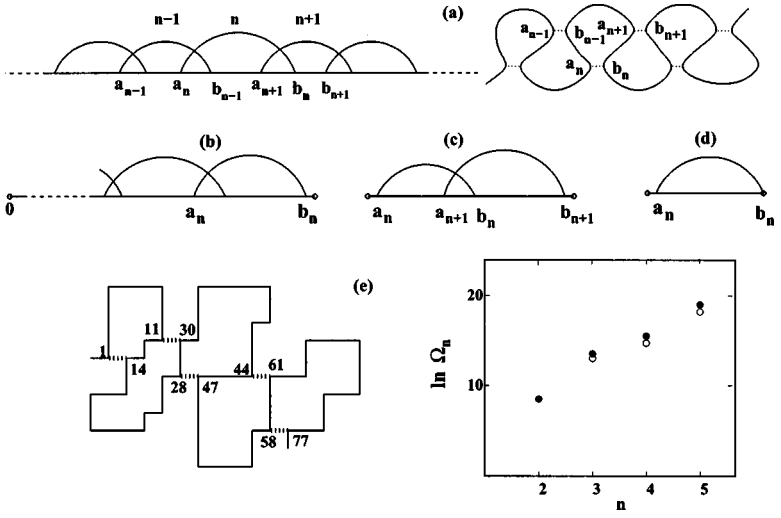

FIG. 5. (a) Multiple crossing links in series, the graph and the conformation. The conformational count for graph in (b) with n crossing links can be computed using conformational counts for subgraphs (c) and (d). (e) Test against exact computer enumeration (filled circles) shows that the theory (empty circles) gives the good estimation of the number of conformations for graphs with crossing links in series.

We can further generalize the above method to treat more complex crossing-linked secondary structures, for example, the graph and structure shown in Fig. 4(a), where RNA secondary structures are brought into contact through the crossing-linked loops A_0 and B_0. For such complex graphs, we need to replace the $S_{A_1}^{(\nu)}$ vector in Eq. (10) for the loop A_1 by the corresponding vector for the complex secondary structure attached to A_0. The computation of such vector for an arbitrary secondary structure is quite straightforward with the previously developed matrix method[12] by replacing $S_{A_1}^{(\nu)}$ with a product of matrices for the secondary structural units.

If loop B_0 also has a complex secondary structure attached, similar to Eq. (10), we have

$$\Omega_f(f_B,\mathbf{I}) = \omega_f(y_{\text{int}},l_{\text{int}},f_B)\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}\alpha^{(\mu)}Y_{\mu\nu}S_{A_1}^{(\nu)}, \tag{11}$$

where $S_{A_1}^{(\nu)}$ is for the secondary structure attached to B_0. With Eqs. (10)–(12), we can compute the number of conformations for any crossing-linked arbitrary secondary structures.

C. Graphs with multiple crossing links in series

We can apply the above approach to treat graphs with a series of crossing links; see Fig. 5(a). We use Ω_n, $\omega_2(n)$, and $\omega_1(n)$ $(n=1,2,\ldots)$ to denote the number of conformations for the graph in Figs. 5(b) (with n crossing links), 5(c) and 5(d), respectively. Using the following approximation:

$$\frac{\Omega_{n+1}}{\Omega_n} \simeq \frac{\omega_2(n)}{\omega_1(n)}, \tag{12}$$

we have

$$\Omega_n \simeq \Omega_2\prod_{r=2}^{n-1}\frac{\omega_2(r)}{\omega_1(r)}. \tag{13}$$

As shown in Fig. 5(e), test against exact computer enumeration in two-dimensional lattice model shows that Eq. (5) can give a good estimation for Ω_n.

IV. GRAPHS WITH MULTIPLE CROSSING LINKS

A. Graphs with a tertiary contact added to a set of nested contacts

In this section we treat graphs which contain, as shown in Fig. 6(a), a tertiary contact $(j,N-1)$ that crosses two nested links $(1,i)$ and (m,k). In contrast to the graph in Fig. 1, an additional contact is established between monomers m and k in Fig. 6(a), resulting in two nested loops A_1 and A_2.

To account for the conformational constraint arising from the contact (k,m), we include the conformation of the (k,m) contact in the enlarged interface $\mathbf{I}$; see the thick lines in Fig. 6(a). Correspondingly, we define the free loop segments F_{A_1}, F_{A_2}, and F_B as the chain segments from monomer 2 to monomer $m-1$, from $m+1$ to $j-1$, and from $i+1$ to $N-2$, respectively. The sum over all the possible conformations of $\mathbf{I}$ gives the number of conformations for the graph:

$$\Omega = \sum_I \Omega_f(f_{A_1},\mathbf{I})\Omega_f(f_{A_2},\mathbf{I})\Omega_f(f_B,\mathbf{I}), \tag{14}$$

where $\Omega_f(f_{A_1},\mathbf{I})$, $\Omega_f(f_{A_2},\mathbf{I})$, and $\Omega_f(f_B,\mathbf{I})$ are the numbers of conformations of the respective free loop segments, and can be given by the ω_f function defined in Eq. (4) and tabulated in Table I:

$$\Omega_f(f_{A_1},\mathbf{I}) = \omega_f(y_{\text{int}}^{(A_1)},l_{\text{int}}^{(A_1)},f_{A_1}) = \omega_f(y_k-y_i,i-k,m-3), \tag{15}$$

$$\Omega_f(f_{A_2},\mathbf{I}) = \omega_f(y_{\text{int}}^{(A_2)},l_{\text{int}}^{(A_2)},f_{A_2}) = \omega_f(y_j-y_k,k-j,j-m-2), \tag{16}$$

$$\Omega_f(f_B,\mathbf{I}) = \omega_f(y_{\text{int}}^{(B)},l_{\text{int}}^{(B)},f_B) = \omega_f(y_i-y_j,i-j,N-i-3). \tag{17}$$

Using the above three equations, for a given enlarge interface conformation $\mathbf{I}$, we can obtain $\Omega_f(f_{A_1},\mathbf{I})$, $\Omega_f(f_{A_2},\mathbf{I})$, and $\Omega_f(f_B,\mathbf{I})$ from Table I. Figure 6(b) shows that the method for the calculations of Ω is reliable as tested against the exact computer enumeration.

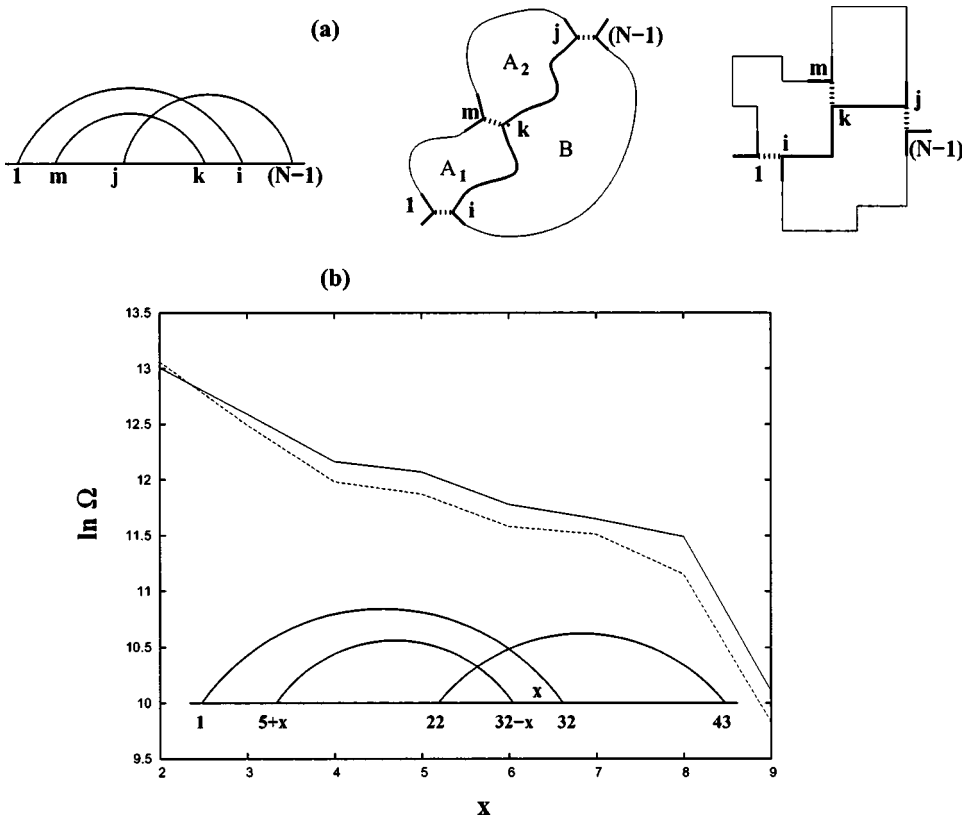

FIG. 6. (a) The conformation with three crossing links and the corresponding graph. The additional contact is established between monomer of free loop segment *m* and an interfacial monomer *k*. The thick lines denote the enlarged interface **I**. (b) The test of the theory (dashed line) against exact enumeration (solid line) using the graph with variable position of the middle contact.

We can generalize the above approach to treat graphs with more crossing links; see Fig. 7(a). Similar to Eq. (14), we have

$$\Omega = \sum_{I} \left(\prod_{r=1}^{n} \Omega_f(f_{A_r}, \mathbf{I}) \right) \Omega_f(f_B, \mathbf{I}), \tag{18}$$

where **I** is the conformation of the enlarged interface [shown as thick lines in Fig. 7(a)], and $\Omega_f(f_{A_r}, \mathbf{I})$ $(r=1,2,\ldots,n)$ and $\Omega_f(f_B, \mathbf{I})$ are the numbers of conformations of the free loop segments in loops A_r and B [see Fig. 7(a)]. Using Eqs. (15)–(23), for each given **I**, we can calculate $\Omega_f(f_{A_r}, \mathbf{I})$ $(r=1,2,\ldots,n)$ and $\Omega_f(f_B, \mathbf{I})$ in terms of the ω_f function tabulated in Table I.

B. Graphs with a tertiary contact added to a secondary structure

Using Eqs. (19) and (20), we can treat more complex graphs with complex secondary structures attached to loops

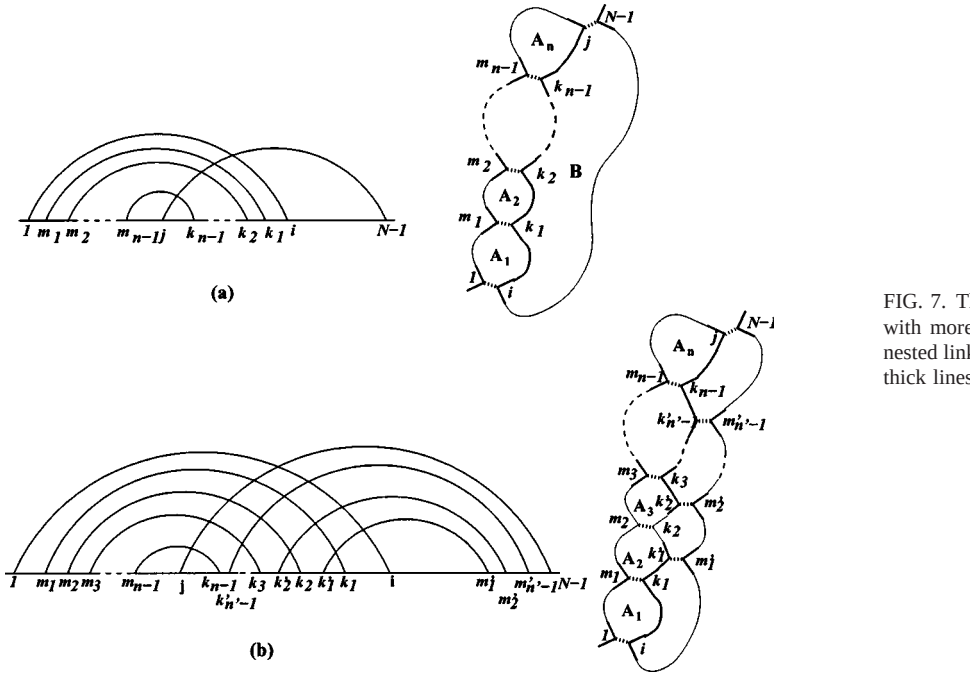

FIG. 7. The method can be generalized to treat graphs with more crossing links (a) and two crossing sets of nested links (b). The enlarged interfaces are shown with thick lines.

A_r $(r=1,2,\cdots,n)$ and B in Fig. 7(a). To account for the conformations of the secondary structures attached to the loops, we use Eq. (10) to calculate $\Omega_f(f_{A_r},\mathbf{I})$ and $\Omega_f(f_B,\mathbf{I})$ in Eq. (18):

$$\Omega_f(f_{A_r},\mathbf{I}) = \omega_f(y_{\text{int}}^{(A_r)},l_{\text{int}}^{(A_r)},f_{A_r})\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}\alpha^{(\mu)}Y_{\mu\nu}S_{A_r}^{(\nu)}, \tag{19}$$

$$\Omega_f(l,B) = \omega_f(y_{\text{int}}^{(B)},l_{\text{int}}^{(B)},f_B)\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}\alpha^{(\mu)}Y_{\mu\nu}S_{B}^{(\nu)}, \tag{20}$$

where $S_{A_r}^{(\nu)}$ and $S_B^{(\nu)}$ are the conformational counts for the secondary structures attached to A_r and to B, respectively. Substituting the above results for $\Omega_f(f_{A_r},\mathbf{I})$ and $\Omega_f(f_B,\mathbf{I})$ in Eq. (18) yields the number of conformations Ω for the graph.

C. Multiple crossing links between nested contacts

The graph in Fig. 7(b) consists of two crossing-linked sets of nested links. The enlarged interface $\mathbf{I}$ is shown as thick lines in Fig. 7(b). The sum over all the possible conformations for the enlarged interface gives the number of conformations for the graph:

$$\Omega = \sum_{I}\prod_{r=1}^{n}\Omega_f(f_{A_r},\mathbf{I})\prod_{s=1}^{n'}\Omega_f(B_s,\mathbf{I}), \tag{21}$$

where F_{A_r} and F_{B_s} are the free loop segments from monomer $m_{r-1}+1$ to monomer m_r-1 in loop A_r and from monomer $m'_{s-1}+1$ to monomer m'_s-1 in loop B_s, respectively. Here monomers $1,i,j,N-1$ are regarded as m_0,k_0,k'_n,m'_n, respectively. $\Omega_f(f_{A_r},\mathbf{I})$ and $\Omega_f(B_s,\mathbf{I})$ are the numbers of conformations of F_{A_r} and F_{B_s}, which can be obtained through the ω_f function, as shown in Eqs. (15), (16), and (23):

$$\Omega_f(f_{A_r},\mathbf{I}) = \omega_f(y_{\text{int}}^{(A_r)},l_{\text{int}}^{(A_r)},f_{A_r}), \tag{22}$$

$$\Omega_f(B_s,\mathbf{I}) = \omega_f(y_{\text{int}}^{(B_s)},l_{\text{int}}^{(B_s)},f_{B_s}), \tag{23}$$

where $y_{\text{int}}^{(A_r)}$ is the y component of the end-end vector $k_{r-1}\rightarrow k_r$ for the interface from monomer k_{r-1} to monomer k_r, $l_{\text{int}}^{(A_r)}=k_{r-1}-k_r$ is the chain length of the interface, and $f_{A_r}=m_r-m_{r-1}-2$ is the chain length of F_{A_r}; and $y_{\text{int}}^{(B_s)}$ is the y component of the end-end vector $k'_s\rightarrow k'_{s-1}$ for the interface from monomer k'_{s-1} to monomer k'_s, $l_{\text{int}}^{(B_s)}=k'_{s-1}-l'_s$ is the chain length of the interface, and $f_{B_s}=m'_s-m'_{s-1}-2$ is the chain length of F_{B_s}.

V. ILLUSTRATIVE CALCULATION FOR THE PARTITION FUNCTION

The partition function is defined in Eq. (1) as a sum over all the possible graphs. Therefore, the first step toward its calculation is to enumerate graphs. We will enumerate all the graphs involving up to four crossing links formed by a tertiary contact. We assign the interaction energy $-\epsilon_2$ for each secondary contact and $-\epsilon_3$ for each tertiary contact. ϵ_3 can be different from ϵ_2. The energy of a graph is equal to $-\epsilon_2$ [the number of secondary contacts] $-\epsilon_3$ [the number of tertiary contacts]. Though the present theory can treat the sequence dependence of the chain, for the purpose of an illustrative calculation, here we do not take into account the sequence and temperature dependence of ϵ_2 and ϵ_3. Effectively, we consider a homopolymer. To simplify the calculation, we consider relatively short chains of less than 35 monomers (nucleotides). For longer chains, we need to include more complex graphs with more tertiary contacts and crossing links.

A. Secondary and tertiary structure elements

We can classify the crossing-linked graphs into three groups according to the number of the crossing links (two, three, and four); see Fig. 8(a) for representative examples for each group of the graphs. For a given chain length, we exhaustively enumerate all the possible arrangements of the crossing links and shuffle secondary links around all possible positions on the graph so that the total number of links do not exceed four. For each generated graph, we compute the number of accessible chain conformations Ω. Figure 8(b) shows the total number of (two-dimensional lattice) conformations for each group of the graphs for different chain lengths. Also plotted in this figure is the result from the exact computer enumeration. We find good agreement between the two sets of results, especially for two- and three-crossing links graphs, the theory and the computer enumeration give nearly identical results.

The graphs so far considered do not contain tails, and we call them the structure elements. Chains which are not longer than 35 nucleotides and have up to four contacts can fold into a larger number of different secondary and tertiary structure elements. A secondary structure from monomer a to monomer b is an element if an outermost contact $(a+1,b-1)$ is formed. For a chain segment between monomers a and b, we enumerate all the possible tertiary and secondary elements. For each tertiary element, we use the theory developed above to compute the number of chain conformations and to calculate the energy in terms of $-\epsilon_2$ and $-\epsilon_3$. From Eq. (1), the sum over all the possible elements gives the "element" partition function for the segment between a and b. For a homopolymer with given ϵ_2 and ϵ_3, such partition function is a function of the length $l=b-a$ only. So we denote the element partition function as $Q_0(l)$.

B. Graphs consisting of one or more structure elements and single-stranded segments

To obtain the full partition function of the chain, we need to add the contributions from the tails. For a given chain length L, the graph generally can contain multiple tertiary or secondary structure elements and the tails. To calculate the number of conformations for such structures, one can, roughly speaking, multiply the numbers of conformations of each element and of each single-stranded segment.

Graphs with a single secondary or tertiary structure element. The partition function of one (tertiary or secondary) structure element with tails is given by $Q_0(l)\omega_t(t_1)\omega_t(t_2)$, where $\omega_t(t_i)$ is the number of conformations of the tail of length t_i (see Table III). In fact, the number of tail conforma-

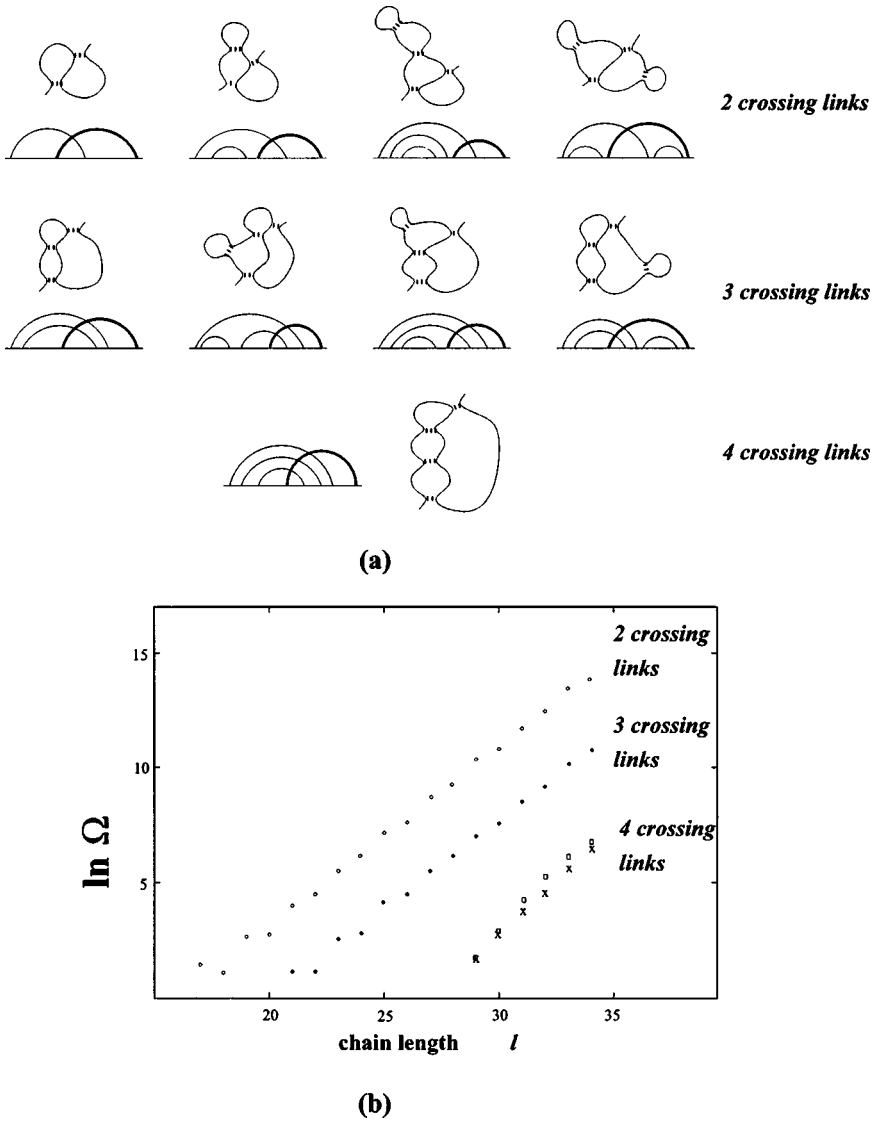

FIG. 8. (a) The three groups of the graphs (with two, three, and four crossing links) and their representative conformations. The tertiary contact is shown in bold. The structures presented are called tertiary structure elements because they do not contain tails. (b) The total number of conformations of each group as a function of the chain length l. The total number of contacts is restricted to be less than or equal to four. The results of approximate and exact calculations are not distinguishable for two and three crossing linked graphs. For the case of four crossing links, squares correspond to exact calculations, and crosses represent results from our theory.

tions depends mainly on the total length of tails $t=t_1+t_2$, and as an approximation, we have $\omega_t(t_1)\omega_t(t_2)\simeq\omega_t(t-1)$. For a given total chain length L, there are $t-1=L-l+1$ possible positions for a single structure element of chain length l. Therefore, the total partition function for all such graphs is

$$Q_1 \simeq Q_0(l)\omega_t(L-l+1)(L-l+1). \tag{24}$$

Graphs with two structure elements. In this case, the chain has three single-stranded segments, each with chain length denoted by t_1, t_2, and t_3. The total length of the single-stranded segments can be determined from the total chain length L and lengths l_1, l_2 of each structure element: $t=t_1+t_2+t_3=L-l_1-l_2+4$, and the partition function can be approximated by $Q_0(l_1)Q_0(l_2)\omega_t(L-l_1-l_2+1)$. The number of graphs containing two such elements and single-stranded segments of total length t can be found from the following considerations. t_1 can have $t-2$ possible values: $t_1=1,2,\ldots,(t-2)$, t_2 can be chosen from $(t-t_1-1)$ possible values, and $t_3=t-t_1-t_2$. Therefore, the number of graphs for given l_1 and l_2 is $\Sigma_{i=2}^{t-1}(t-i)=(t-2)(t-1)/2$, and the sum over all such graphs gives the partition function for graphs with two structure elements:

$$Q_2 \simeq \tfrac{1}{2}(L-l_1-l_2+2)(L-l_1-l_2+3)Q_0(l_1) \times Q_0(l_2)\omega_t(L-l_1-l_2+1). \tag{25}$$

Graphs with three structure elements. The maximum number of structures in sequence which is allowed by restriction imposed on the total chain length is three. We estimated the corresponding partition function by

TABLE III. The number of tail conformations ω_t as a function of the tail length t calculated by exact computer enumeration on two-dimensional lattice.

t	1	2	3	4	5	6	7	8	9	10
ω_t	1	2	4	9	21	50	118	281	666	1 584
t	11	12	13	14	15	16	17	18	19	20
ω_t	3743	8877	20 934	49 522	116 579	275 205	646 909	1 524 458	3 579 101	8 418 185

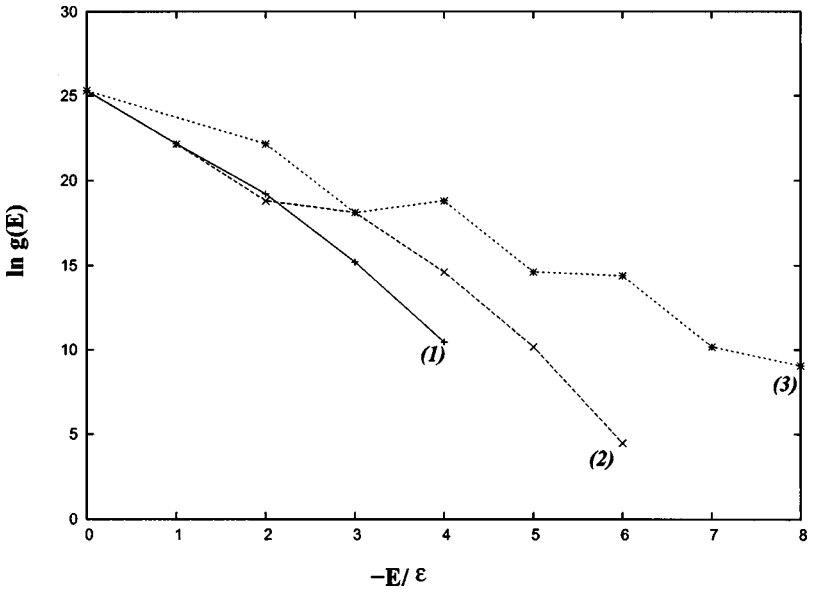

FIG. 9. The density of states of a 32-mer homopolymer calculated for $(\epsilon_2,\epsilon_3)=(\epsilon,\epsilon)$ for line (1), $(\epsilon,2\epsilon)$ for line (2), $(2\epsilon,\epsilon)$ for line (3).

$$Q_3 \simeq Q_0(l_1)Q_0(l_2)Q_0(l_3)\omega_t(L-l_1-l_2-l_3+1) \times \sum_{i=1}^{L-l_1-l_2-l_3+3} i(i+1)/2, \tag{26}$$

where the coefficient $\Sigma_{i=1}^{L-l_1-l_2-l_3+3} i(i+1)/2$ is the number of possible sets (t_1,t_2,t_3,t_4) obtained from considerations similar to those described above.

C. Density of states and partition function

To compute the density of states $g(E)$ (= the total number of all the conformations that have energy E) for the chain of the given length, we consider all the secondary and tertiary structure elements along with the tails, such that the total chain length is fixed and the total number of contacts does not exceed four. The result for a 32-mer homopolymer is shown in Fig. 9 for three different values of (ϵ_2,ϵ_3).

From $g(E)$, we can compute the partition function as a sum over possible energy values:

$$Q(T)=\sum_E g(E)e^{-E/k_BT}. \tag{27}$$

From the partition function $Q(T)$, we can compute the heat capacity $C(T)=\partial/\partial T[k_BT^2(\partial/\partial T)\ln Q]$. The temperature dependence of the heat capacity (the melting curve) for a 30 mer homopolymer and $\epsilon_2=\epsilon_3=\epsilon$ shows a single transition at the melting temperature $k_BT_m/\epsilon=0.22$. The energy dependence of the microcanonical ensemble free energy $F(E)=E-k_BT\ln g(E)$ for different temperatures is plotted in Fig. 10. In order to separate conformations with and without tertiary interactions, we show them as separate points shifted along the x axis by 0, 0.1, and 0.2 for conformations with zero, one, and two tertiary contact(s). For example, for $E=-4\epsilon$, we have three sets of points, corresponding to conformations that have (four secondary contacts), (three secondary and one tertiary contacts), and (two secondary and two tertiary contacts), respectively; see Fig. 10. We define a contact as a secondary structural contact if it is part of a secondary

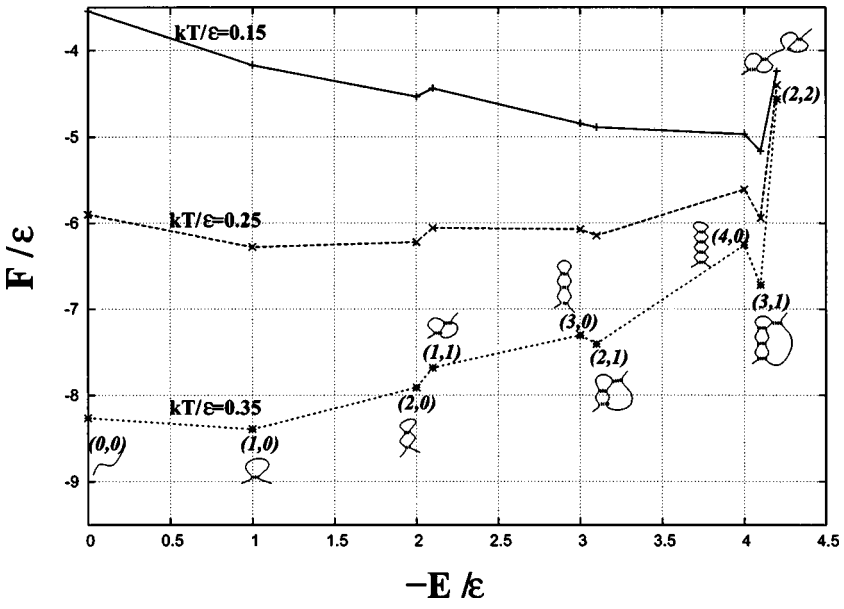

FIG. 10. The free energy as a function of the energy for different k_BT and $\epsilon_2=\epsilon_3=\epsilon$. Since the energies of secondary and tertiary contacts are taken to be equal, the secondary and tertiary conformations would be indistinguishable in the plot. To separate them, the free energies of conformations with one and two tertiary contact(s) are shown by points shifted by 0.1 and 0.2 to the right along the x axis, respectively. The points on the graph are marked (i,j), where i is the number of secondary contacts and j is the number of tertiary contacts. For each point, a representative conformation is shown. For example, states (3,0) and (2,1) have the same energy 3ϵ, but with 0 and 1 tertiary contact, respectively.

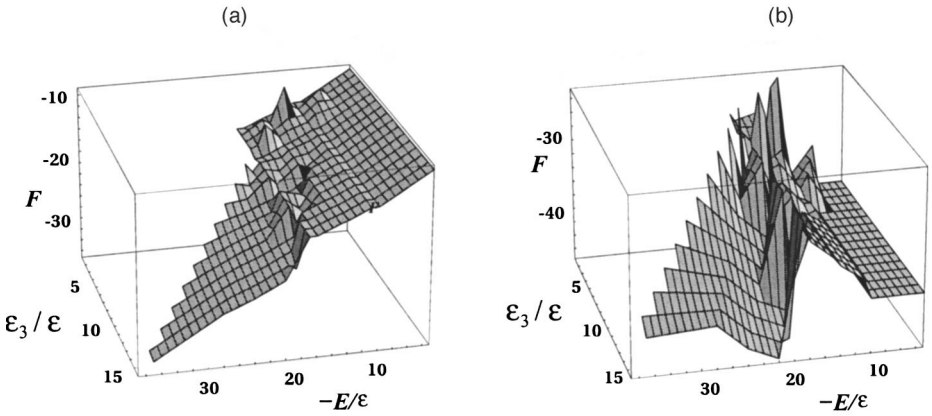

FIG. 11. The free energy landscape at temperatures $k_BT/\epsilon=0.4$ (a) and $k_BT/\epsilon=1.8$ (b) for a 30-mer homopolymer with $\epsilon_2=4\epsilon$ and $0\leqslant\epsilon_3\leqslant 15\epsilon$.

structure (= set of nested or unrelated contacts), and a tertiary contact if breaking it would cause the crossing-linked (tertiary) conformation become a secondary structure. For some simple conformations, the distinction between secondary and tertiary contacts is not unambiguous. For example, in the conformations marked (1, 1) in Fig. 10, both contacts can be either secondary or tertiary. The free energy plot reveals that the most stable state at low temperature, i.e., the lowest energy state, is the one with three secondary structural contacts and one tertiary contact [(3,1) in Fig. 10]. The unfolding transition from this "native" state to an ensemble of unfolded and partially unfolded states occurs around T_m when they have about the equal free energies.

VI. INTERPLAY BETWEEN SECONDARY AND TERTIARY INTERACTIONS

In order to study the interplay between the secondary and tertiary contact energies, we fix the secondary energy parameter $\epsilon_2=4\epsilon$ and change the energy of the tertiary contact ϵ_3. The native structure and the melting temperature will obviously change with the changing of ϵ_3. Figure 11 shows the free energy landscapes F as a function of E and ϵ_3 for a 30-mer chain at different temperatures. We find that at low temperature T, independent of the tertiary energy parameter $\epsilon_3\leqslant 15\epsilon$, the native state (minimum free energy) is always the state with the lowest energy E. As the temperature is increased, there exists a critical tertiary energy parameter ϵ_3^* such that the minimum free energy state is the highest E unfolded state for $\epsilon_3<\epsilon_3^*$ and shifts to a low E state for $\epsilon_3>\epsilon_3^*$. As the tertiary interaction is strengthened, the folding-unfolding melting temperature $T_m(\epsilon_3)$ would increase. So for a given temperature T, there exists a critical ϵ_3^* determined from $T=T_m(\epsilon_3^*)$. For $\epsilon_3<\epsilon_3^*$, $T_m(\epsilon_3)<T$, so the chain is in the unfolded state, and for $\epsilon_3>\epsilon_3^*$, $T_m(\epsilon_3)>T$, so the chain is in the folded state.

For a given set of the (ϵ_2,ϵ_3) parameter, we can calculate the heat capacity melting curve from the partition function. From the melting curve, we can identify the temperatures at which the conformational transitions occur. From the transition temperatures, we divide the temperature range into several pretransition and posttransition regimes. In each regime, we find the most stable state. By performing the analysis for the melting curves and the free energy landscapes for different (ϵ_2,ϵ_3) parameter sets, we are able to obtain the phase diagram for different parameters (T,ϵ_3) (Fig. 12). The chain has different stable structures in different (T,ϵ_3) regions. The stable state in the phase diagram is marked with (n,m) for structures with n secondary and m tertiary contacts. The energy of the corresponding structure is $n\epsilon_2+m\epsilon_3$.

From the phase diagram, we find in the region where ϵ_3 is comparable with ϵ_2, the chain undergoes multiple transitions in the melting process. Overall speaking, the melting is less cooperative due to the interplay between the secondary and tertiary interactions when ϵ_3 is comparable with ϵ_2, and more cooperative when $\epsilon_3\gg\epsilon_2$ or $\epsilon_3\ll\epsilon_2$. In the $\epsilon_3\ll\epsilon_2$ limit, the melting involves the secondary structural changes only, such as $(4,0)\rightarrow(3,0)\rightarrow(2,0)\rightarrow(1,0)\rightarrow(0,0)$ for $\epsilon_3=1$, here (m,n) denotes states with m secondary and n tertiary contacts. In the $\epsilon_3\gg\epsilon_2$ limit, the melting transitions mainly involve the breaking of the tertiary contacts [e.g., $(2,2)\rightarrow(1,1)\rightarrow(0,0)$ for $\epsilon_3=14$]. In the intermediate range of ϵ_3, the melting transitions involve the change of either the secondary or the tertiary structural contacts [e.g., $(3,1)\rightarrow(2,1)\rightarrow(1,1)\rightarrow(1,0)\rightarrow(0,0)$ for $\epsilon_3=7$].

VII. SUMMARY

In the present study, we have established a statistical mechanical machinery for simple RNA tertiary contacts to treat the nonadditive chain entropy and the partition function, from which the thermodynamic properties can be predicted. The method that we have developed enables the calculation for the number of chain conformations and the partition function of the RNA-like molecules with simple tertiary interactions. The key idea of the method is to use the intrachain contacts to subdivide the conformation into different loops, and to assume that the excluded volume interferences between the loops predominantly come from the excluded volume of the monomers near the interfaces between the loops. The method has been shown to give accurate results for two-dimensional lattice test systems. Several simple types of tertiary folds are considered in the present work. These types of tertiary folds represent a large class of RNA tertiary structures. Applications to an illustrative simple model suggest that the interplay between the secondary and tertiary interactions can cause rugged free energy landscape and noncooperative melting transitions. Moreover, the generality of the above basic idea for the method suggests that the method may be extended to treat more complex tertiary topologies

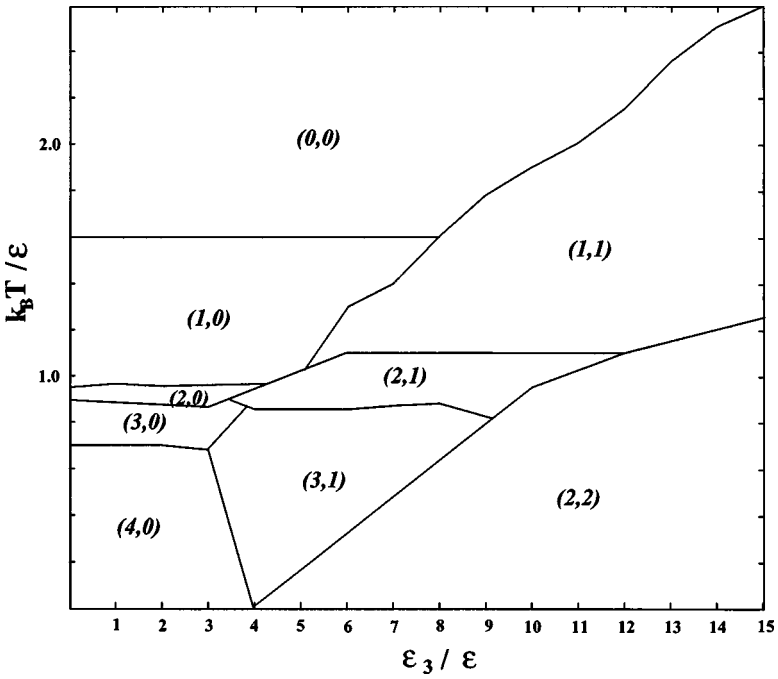

FIG. 12. The phase diagram for a 30-mer chain. (m,n) denotes conformations with m secondary and n tertiary contacts.

that involve multiple crossing-linked tertiary contacts. In addition, the method is developed based on the graphic representation of the structure and is thus general in terms of chain representation. The method can be implemented in more realistic off-lattice chain representations. Tertiary structure thermal stability is strongly dependent on the ionic solution condition. The electrostatic effect is not the focus here. But the model developed here would provide a more complete statistical mechanical framework for the modeling of the electrostatic interactions.

ACKNOWLEDGMENTS

The authors thank Dr. Wenbing Zhang for useful discussions. This research was supported by National Institutes of Health (NIH) (Grant No. GM063732) (to S.J.C.).

[1]M. J. Serra and D. H. Turner, Methods Enzymol. **259**, 242 (1995).
[2]D. H. Turner, N. Sugimoto, and S. M. Freier, Annu. Rev. Biophys. Biophys. Chem. **17**, 167 (1998).
[3]J. A. Jaeger, J. SantaLucia, Jr., and I. Tinoco, Jr., Annu. Rev. Biochem. **62**, 255 (1993).
[4]M. Zuker, Science **224**, 48 (1989).
[5]J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[6]S. M. Freier *et al.*, Proc. Natl. Acad. Sci. U.S.A. **83**, 9373 (1986).
[7]J. A. Jaeger, D. H. Turner, and M. Zuker, Proc. Natl. Acad. Sci. U.S.A. **86**, 7706 (1989).
[8]A. E. Walter *et al.*, Proc. Natl. Acad. Sci. U.S.A. **91**, 9218 (1994).
[9]R. Nussinov and A. B. Jacobson, Proc. Natl. Acad. Sci. U.S.A. **77**, 6309 (1980).
[10]M. Zuker and P. Stiegler, Nucleic Acids Res. **9**, 133 (1981).
[11]M. Zuker and D. Sankoff, Bull. Math. Biol. **46**, 591 (1984).
[12]S.-J. Chen and K. A. Dill, J. Chem. Phys. **103**, 5802 (1995).
[13]S.-J. Chen and K. A. Dill, J. Chem. Phys. **109**, 4602 (1998).
[14]S.-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **97**, 646 (2000).
[15]W. Zhang and S.-J. Chen, J. Chem. Phys. **114**, 7669 (2001).
[16]W. Zhang and S.-J. Chen, J. Chem. Phys. **114**, 7669 (2001).
[17]R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[18]D. Thirumalai, N. Lee, S. A. Woodson, and D. K. Klimov, Annu. Rev. Phys. Chem. **52**, 751 (2001).
[19]T. C. Gluick, N. M. Wills, R. F. Gesteland, and D. E. Draper, Biochemistry **36**, 16173 (1997).
[20]K. S. Silverman, M. X. Zheng, M. Wu, I. Tinoco, Jr., and T. R. Cech, RNA **5**, 1665 (1999).
[21]R. Das *et al.*, J. Mol. Biol. **332**, 311 (2003).
[22]X. Fang, T. Pan, and T. R. Sosnick, Biochemistry **38**, 16840 (1999).
[23]S. L. Heilman-Miller, D. Thirumalai, and S. A. Woodson, J. Mol. Biol. **306**, 1157 (2001).
[24]S. L. Heilman-Miller, J. Pan, D. Thirumalai, and S. A. Woodson, J. Mol. Biol. **309**, 57 (2001).
[25]M. S. Rook, D. K. Treiber, and J. R. Williamson, Proc. Natl. Acad. Sci. U.S.A. **96**, 12471 (1999).
[26]E. ten Dam, K. Pleij, and D. Draper, Biochemistry **31**, 11665 (1992).
[27]C. W. A. Pleij, Curr. Opin. Struct. Biol. **4**, 337 (1994).
[28]B. A. L. M. Deiman and C. W. A. Pleij, Semin. Virol. **8**, 166 (1997).
[29]E. Rivas and S. R. Eddy, J. Mol. Biol. **285**, 2053 (1999).
[30]J. P. Abrahams, M. van den Berg, E. van Batenburg, and C. Pleij, Nucleic Acids Res. **18**, 3035 (1990).
[31]A. P. Gultyaev, F. H. D. van Batenburg, and C. W. A. Pleij, J. Mol. Biol. **250**, 37 (1995).
[32]H. Isambert and E. D. Siggia, Proc. Natl. Acad. Sci. U.S.A. **97**, 6515 (2000).
[33]C. A. Theimer *et al.*, J. Mol. Biol. **279**, 545 (1998).
[34]C. A. Theimer and D. P. Giedroc, J. Mol. Biol. **289**, 1283 (1999).
[35]A. P. Gultyaev, F. H. D. van Batenburg, and C. W. A. Pleij, RNA **5**, 609 (1999).
[36]A. Lucas and K. A. Dill, J. Chem. Phys. **119**, 2414 (2003).
[37]K. M. Fiebig and K. A. Dill, J. Chem. Phys. **98**, 3475 (1993).