A three-dimensional statistical mechanical model of folding double-stranded chain molecules

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

(Received 5 January 2001; accepted 16 February 2001)

Based on a graphical representation of intrachain contacts, we have developed a new three-dimensional model for the statistical mechanics of double-stranded chain molecules. The theory has been tested and validated for the cubic lattice chain conformations. The statistical mechanical model can be applied to the equilibrium folding thermodynamics of a large class of chain molecules, including protein β-hairpin conformations and RNA secondary structures. The application of a previously developed two-dimensional model to RNA secondary structure folding thermodynamics generally overestimates the breadth of the melting curves [S-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **97**, 646 (2000)], suggesting an underestimation for the sharpness of the conformational transitions. In this work, we show that the new three-dimensional model gives much sharper melting curves than the two-dimensional model. We believe that the new three-dimensional model may give much improved predictions for the thermodynamic properties of RNA conformational changes than the previous two-dimensional model. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1363670]

I. INTRODUCTION

The thermodynamic properties of chain molecules are governed by the statistical mechanics of the conformational ensembles. Computation of the statistical thermodynamics of polymer chains, including the conformational distributions, the partition functions, and the free energies, poses significant and challenging problems. Computer simulational approaches can accurately treat the atomic details of the chains, but often have the problems with conformational sampling, especially for large conformational transitions.[1,2] Simplified lattice models give rigorous account for the principles for the conformational statistics of chains molecules, but in general lack the ability to give predictions for specific sequences and for chains of realistic length.[3–10] Polymer theory approach, e.g., the rotational isomeric states (RIS) models,[11–15] can compute the chain conformational statistics for long chains with atomic detail, but are limited by the inability to treat nonlocal intrachain interactions that often dictate the folding of compact structures of biopolymers.

Recently, a new computational model[16–18] has been developed for a large class of chain conformations—double-stranded chain conformations, e.g., RNA secondary structures[19,20] and two-stranded antiparallel β-sheet in proteins. The new model accurately accounts for the sequence dependence, nonlocal intrachain contacts, and the excluded volume interactions. Excluded volume accounts for the effect that no two monomers can occupy the same position in space. Applications of the model gave good results in predicting melting thermodynamics for RNA secondary structures.[18] However, the accuracy of the predictions is limited by, among other factors, the simplified two-dimensional (2D) representation of chain conformations. One of the problems that may be caused by the two-dimensional chain representation is that the predicted thermal melting curves are in general wider than the experimental results,[18] suggesting an underestimation for the sharpness of the structural transitions. An accurate prediction for the nature of the conformational transition for realistic chain molecules requires a three-dimensional (3D) model. Our purpose here is to move beyond the two-dimensional model to explore a three-dimensional statistical mechanical model, and show that the three-dimensional model predicts much sharper conformational transitions than the two-dimensional model.

II. THE STATISTICAL MECHANICAL MODEL

At the center of the statistical thermodynamics is the partition function $Q(T)$,

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

where T is the temperature, k_B is the Boltzmann constant, Σ_{conf} is the sum over all the possible conformations, and E is the energy of the conformation. Two ingredients that determine the accuracy of the statistical mechanical model are the energy E for a given conformation and the count of the accessible conformations (Σ_{conf}).

We first define the graphical representation used in our theory for the chain conformations. We model the polymer chain as a sequence of covalently connected monomers positioned on a cubic lattice. We describe a chain conformation by the intrachain contacts—pairs of adjacent but not co-

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

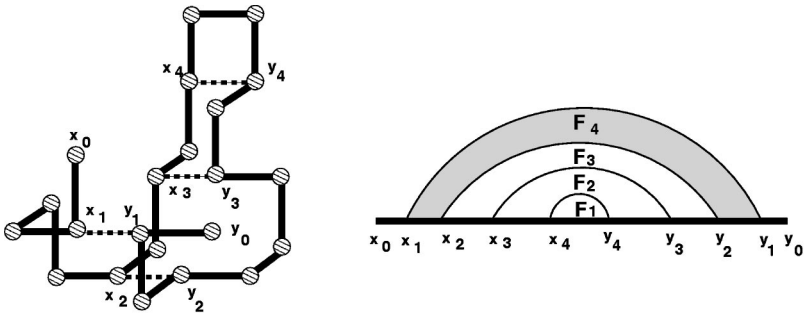

FIG. 1. A nested polymer graph and the corresponding 3D cubic lattice chain conformation. F_1, F_2, F_3, and F_4 are the subunits of the graph. The thick lines represent the covalently bonded polymeric chain, and curved links in the polymer graph and the dashed lines in the 3D lattice conformation represent the intrachain contacts.

valently bonded monomers on the lattice. Mathematically, in our model, we use the a graphical representation—the *polymer graph*[16] to describe the intrachain contacts. As shown in Fig. 1, in a polymer graph, the chain backbone is represented by a straight line, and the intrachain spatial contacts are represented by curved links. Any two curved links bear three type of relationships: nested, unrelated, and (crossing) linked (see Fig. 2). Obviously, a polymer graph generally corresponds to a large number of accessible chain conformations that satisfy the given intrachain contacts.

In terms of the polymer graph, the partition function can be computed as the sum of the statistical weight for all the possible polymer graphs,

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

where Ω is the number of accessible chain conformations for a given graph, and E is the energy of the conformations. In Eq. (1), we assume that energy E is determined by the intrachain contacts, and hence all conformations corresponding to a given polymer graph have the same energy E and the same Boltzmann statistical weight e^{-E/k_BT}.

As a first step toward more complex chain conformations, in this work we focus on the *double-stranded chain conformations*. In the graphical representation, double-stranded chain conformations can be defined by polymer graphs that do not contain crossing links. In Fig. 3, we show a general double-stranded chain conformation and the corresponding polymer graph. Examples of double-stranded chain conformations include RNA secondary structures, two-stranded protein antiparallel β strands.

The basic idea of our approach to the calculation of the partition function $Q(T)$ is to first collect up chain conformations according to the graphs, for each polymer graph, use an analytical formula to compute the number of accessible chain conformations [Ω in Eq. (1)], and then sum over all the possible graphs through a recursive dynamic algorithm. The great advantage of our theory is that the analytical calculation for Ω and the dynamic algorithm are both computationally efficient, so exhaustive sampling for the ensemble of all the possible double-stranded chain conformations becomes computationally viable. In the following, we first develop an analytical method for the calculation of Ω, the number of chain conformations for a given polymer graph.

III. THE COUNT OF CHAIN CONFORMATIONS FOR A GIVEN POLYMER GRAPH

In order to concentrate on the basic principles of the theory, we will use the nested polymer graphs as a working model to illustrate the methodology. As shown in Fig. 1, nested polymer graphs are the graphs that consist of nested links only. Examples of chain conformations for a nested polymer graph include RNA stem-loop structure and protein β-hairpin structure. Since polymer graphs for general double-stranded chain conformations can be transformed into nested graphs through isomorphic transformations,[16] the methodology developed here would also be applicable to general double-stranded conformations.[16,17] Based on the methodology discussed in this section, we will develop a theory for the partition functions of general double-stranded chain molecules in Sec. IV.

To concisely describe the methods and results, we first define some useful notations and terminology for the polymer graph. We will use notation (x,y) to denote the intrachain contact (=curved link on the polymer graph) between monomer x and y, and use notation $\{x,y\}$ to denote the (sub)graph constituted by monomers from x to y along the sequence.

Like any graph, a polymer graph can be decomposed into subgraphs, and subgraphs can be divided into subunits. We define a subunit of the polymer graph as a region enclosed by curved and straight line links and contains no links in its interior. For example, the shaded region in Fig. 1 is a subunit. A subunit corresponds to a flexible region of

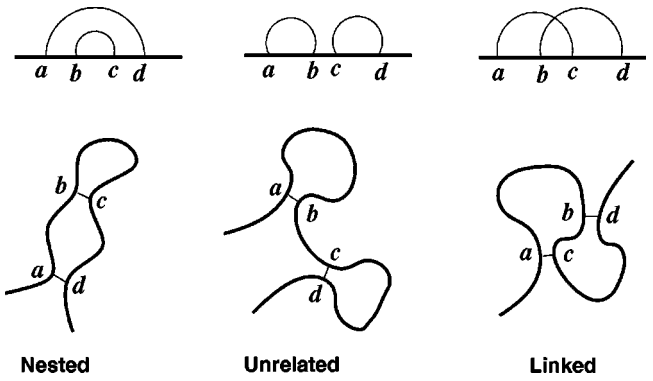

FIG. 2. Three types of relationships between two links of a polymer graph, and the corresponding chain conformations.

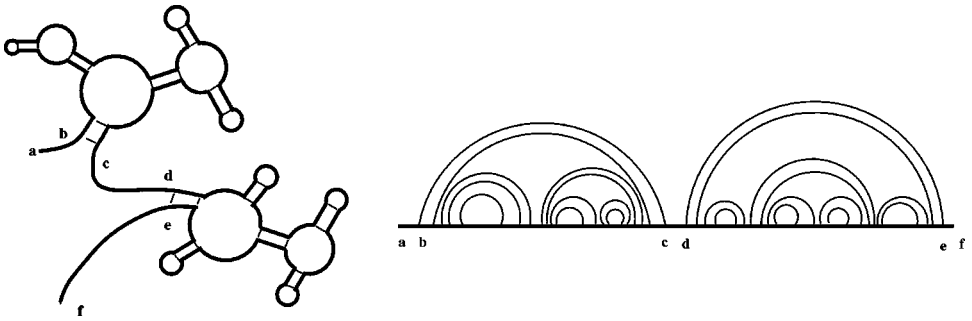

FIG. 3. A schematic general double-stranded chain conformation and the corresponding polymer graph.

doubled-stranded chain, and in general, a subunit corresponds to an ensemble of chain conformations.

To conveniently describe a polymer graph and subgraphs, we classify polymer graphs and subgraphs into two types: ''closed'' and ''open.'' We call a (sub)graph to be ''closed'' if its terminal monomers are connected by a curved link, for example, $\{x_1,y_1\}$ in Fig. 1 is a closed subgraph because it is closed by the curved link (x_1,y_1). We call a (sub)graph to be ''open'' if it does not contain any curved link. Obviously in the conformational space, an open graph corresponds to dangling chain segments that do not involve any intrachain contact, for example, $\{x_0,x_1\}$ and $\{y_1,y_0\}$ in Fig. 1 are open subgraphs. In general, a polymer graph is composed of multiple closed and open subgraphs.

Our calculation for the count of chain conformations for a given polymer graph [Ω in Eq. (1)] involves two steps: we first count chain conformations for the *closed subgraphs*, then obtain Ω for the *whole polymer graph* by including the *open subgraphs*.

A. The count of chain conformation for a closed subgraph

A key component of our theory is the analytical formula for Ω, the number of viable chain conformations accessible to a given graph. A central problem for the count of chain conformations is the excluded volume effect. In the 3D lattice model, the excluded volume is satisfied by the requirement that no any two monomers can occupy the same lattice site. For a given polymer graph, we first divide the polymer graph into *subunits*,[16] compute the conformational count for each subunit rigorously through exact enumeration of self-avoiding walks on cubic lattice, and then ''assemble'' the conformations of the subunits together to obtain the conformations for the whole graph.

1. Classification of subunits

A polymer graph for double-stranded conformations can be factored into subunits, and different subunits are connected through the intrachain contacts. A contact serves as the inlet and outlet for the subunits. For example, subunits F_2 and F_3 in Fig. 1 are connected through contact (x_3,y_3). Therefore, the contact (x_3,y_3) can be treated as the inlet of subunit F_3 and the outlet of subunit F_2. In general, the inlet of a subunit is the outlet of the preceding subunit. Because in a 3D lattice model, each lattice site has coordination number of 6, the maximum (straight line and curved) links connected to a monomer cannot exceed 6. For example, on the polymer graph shown in Fig. 4, monomer A has 7 contacting monomers a_i $(i=1,2,\ldots,7)$, thus the graph is not viable. Therefore, to account for the maximum contact number requirement, we describe the subunits according to the number of curved links connected to the monomers of the inlets and outlets. The number of curved links connected to the inlet and outlet impacts the conformational ensemble available to the subunit upon the connection with other subunits. As shown in Fig. 5, we find that there are 11 types of possible viable subunits of the polymer graph: L, M, R, LR, I, LL, RR, LLR, LRR, LLL, and RRR.

2. The matrix method

a. The subunit conformations: Before considering the connections of the subunit conformations, we first treat chain conformations for each subunit separately. By definition, except at the inlets and outlets, subunit conformations do not contain any intrachain contact, and thus do not involve (sequence-dependent) energy, therefore, in the lattice model, the number of chain conformations and the statistical weight for a subunit is not sequence-dependent. Such property of subunit conformations makes the computation for the subunit partition function simple. Specifically, this property allows us to compute the number of chain conformations for different subunits prior to the calculation for chains with any specific sequence. The conformations of subunits can be counted by the enumeration for the self-avoiding walks on a 3D lattice. By this way, the intra-subunit excluded volume can be treated rigorously. It is important to note that though the sequence effect of the count of subunit conformations is ignored in the lattice model, the energy of any given conformation and the partition function are sequence-dependent.

b. The nearest-neighbor approximation: When subunit conformations are assembled together to form the conformations of the whole graph, the subunit conformations become correlated because some subunit conformations are not vi-

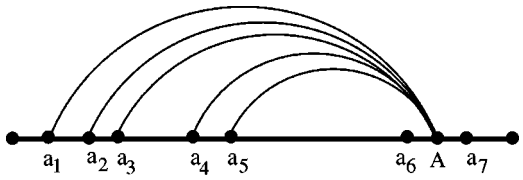

FIG. 4. This graph is not viable because monomer *A* is connected to seven other monomers $(a_i, i=1,2,3,4,5,6,7)$. In a 3D lattice model, each lattice site has six neighboring sites, therefore, considering the two neighboring monomers along the sequence, a monomer can only make no more than four spatial contacts.

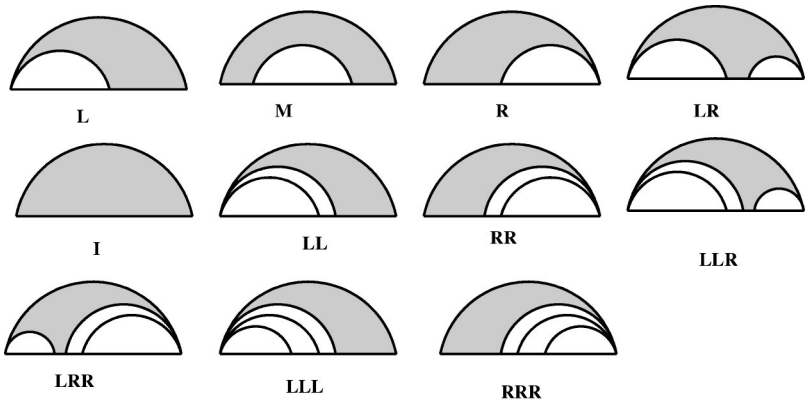

FIG. 5. The 11 types of subunits of a polymer graph and the corresponding representative chain conformations on a cubic lattice.

able due to the excluded volume interactions between subunit conformations. We use the following *nearest-neighbor approximation* to treat the inter-subunit excluded volume interactions. Since monomers are less likely to encounter in space when they are distantly separated along the chain, we consider only the excluded volume interactions between monomers in the vicinity of the inlets and outlets of the nearest neighboring subunits (see type-A interactions shown in Fig. 6). Following the nearest-neighbor approximation, we classify subunit conformations according to the conformations of the inlet and outlet (monomers s and l in Fig. 7) and the nearby monomers (monomers s' and l' in Fig. 7). For 3D cubic lattice chain conformations, we find that the subunit conformations can be classified into six types according to the conformation of the inlet and outlet; see Fig. 7.

c. The S-matrix: According to the inlet and outlet conformations, we define the S-matrix, such that the matrix element s_{ij} $(i,j=1,2,3,4,5,6)$ is the number of subunit conformations in which the inlet is of type-j and the outlet is of type i (see Fig. 7). For small subunits, we calculate the S-matrix by exact computer enumeration for self-avoiding walks on 3D lattice, therefore the intra-subunit excluded volume effect is treated exactly. For large subunits, we calculate the S-matrix through extrapolation from the results for smaller subunits. For an M-mer $(M>18)$ subunit, the S-matrix is given by

$$\ln s_{ij} \simeq a_{ij} M + b_{ij} \ln M + c_{ij}, \tag{2}$$

where $i,j=1,2,3,4,5,6$, and the coefficients a_{ij}, b_{ij}, and c_{ij} are determined by fitting the results of the S-matrix with $M \leqslant 18$. In Table I, we list a_{ij}, b_{ij}, and c_{ij} for type-L,M,I subunits (see Fig. 5). For example, for large M, the S-matrix for type-M subunit is given by

$$\mathbf{S} \simeq p(M) \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

where $p(M)=1.41277M-1.98415\ln M-8.1778$.

d. The Y-matrix: Next we consider the connection between neighboring subunits. According to the nearest-neighbor approximation, the viability of the connection between subunit conformations is predominantly determined by the compatibility of the inlet-outlet conformation. For example, in Fig. 8, we show that the connection between type-1 and type-2 subunit conformations is viable, but the connection between type-2 and type-4 conformations is prohibited by the excluded volume interaction between monomer s_2' and s_4' in the figure. To mathematically describe the connection viability between the subunits, we define a Y-matrix, of which the matrix element $\tilde{y}_{ij}$ $(i,j=1,2,3,4,5,6)$ is defined as the following:

$$\tilde{y}_{ij} = \text{number of different ways for the connection between a type-}j\text{ and a type-}i\text{ conformation.} \tag{3}$$

For example, in Fig. 9, we show that $\tilde{y}_{12}=\tilde{y}_{21}=3$ because there are three different ways for the connection between type-1 and type-2 subunit conformations, corresponding to three rotational degrees of freedom about the contact (s,l). According to Fig. 8, we have $\tilde{y}_{24}=\tilde{y}_{42}=0$.

For 2D lattice conformations,[16,17] the Y-matrix defined above gives quite accurate treatment for the inter-subunit excluded volume effect. But for 3D lattice conformations, each chain segment is less restricted and can take more conformations because of the larger degrees of freedom. As a result, besides the nearest-neighbor monomers (s' and l' in Figs. 7 and 10), the excluded volume interactions between the next-nearest-neighbors (s'' and l'' in Fig. 10) also play important roles and therefore cannot be neglected. As we will see in

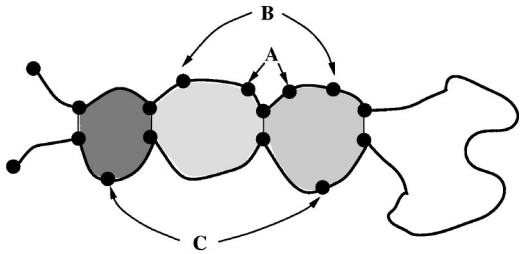

FIG. 6. The excluded volume interaction between subunit conformations: (A) steric interference in the vicinity of the interface between neighboring subunits; (B) steric interference between chain segments away from the interface between neighboring subunits; (C) steric interference between non-nearest-neighbor subunits.

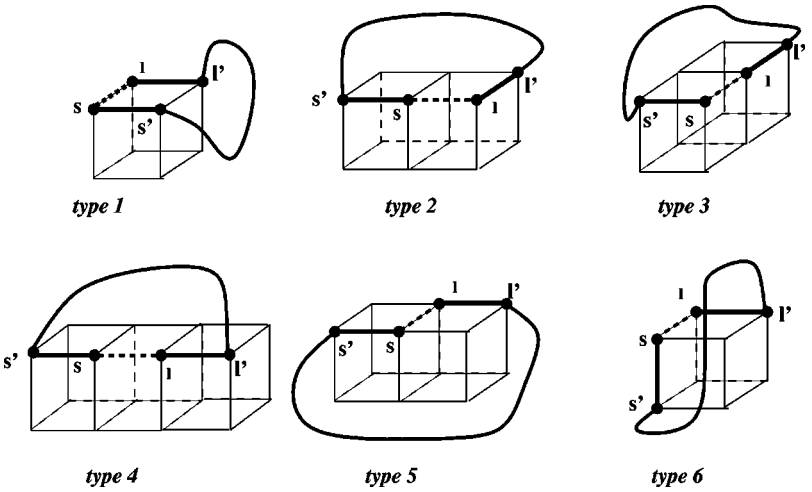

FIG. 7. On a cubic lattice, according to the conformations of the inlet/outlet (s and l) and the nearest neighbor monomers (s' and l'), the subunit conformations can be classified into six types.

Sec. V, the model with Y-matrix defined only by $\tilde{y}_{ij}$, which only accounts for the nearest-neighbor excluded volume, cannot give accurate predictions for the partition function in the 3D model. Therefore, we must go beyond the nearest-neighbor to include the next-nearest-neighbor excluded volume in the Y-matrix.

To account for the next-nearest-neighbor excluded volume, we further classify the subunit conformations according to the next-nearest-neighbors. As an illustration, Fig. 10 shows that type-1 conformations, defined by nearest-neighbor monomers s,s',l,l' shown in Fig. 7, can be further classified into 16 subtypes of conformations according to the positions of the next-nearest-neighbor monomers s'' and l''. Similarly, we found that type-2, 3, 4, 5, and 6 subunit conformations can be classified into 20, 20, 25, 16, and 16 types of conformations according to the next-nearest-neighbor monomers, respectively. Therefore, by including the next-nearest-neighbors, we classify the subunit conformations into 113 $(=16+20+20+25+16+16)$ types.

TABLE I. The asymptotic coefficients a_{ij}, b_{ij}, and c_{ij} for the S-matrix of large subunits defined in Eq. (2). a_{ij}, b_{ij}, and c_{ij} for i and j's not listed in the table are equal to zero. For example, $S_{i1}=S_{1j}=0$ $(i,j=1,2,3,4,5,6)$ for type-M subunits because $a_{ij}=b_{ij}=c_{ij}=0$ for $i=1$ or $j=1$.

Type	ij	a_{ij}	b_{ij}	c_{ij}
I	2 2	1.40983	−1.99	−6.3452
	3 3	1.40983	−1.99	−6.3452
	4 4	1.122	3.46	−18.113
	5 5	1.2078	1.9696	−14.3526
	6 6	0.85	7.2	−22.31
L	2 2	1.47263	−3.15615	−4.37297
	3 3	1.52125	−4.2039	−3.11496
	3 5	0.9364	7.085	−25.6049
	3 6	1.47404	−3.3118	−4.41796
	4 2	1.29522	0.122967	−11.8347
	5 3	0.936402	7.085	−25.6049
	5 5	1.2131	1.9334	−15.8567
	5 6	1.38873	−1.5159	−8.2938
	6 3	1.47404	−3.3118	−4.4178
	6 5	1.38873	−1.559	−8.2938
	6 6	1.4231	−2.179	−6.55534
M	$i,j\neq 1$	1.41277	−1.98415	−8.1778

A straightforward way to account for the next-nearest-neighbor excluded volume is to define an augmented Y-matrix for the 113 types of conformations. But such an approach is computationally not viable because of the large size (113×113) of the **S** and **Y** matrices. Therefore, we introduce an effective 6×6 Y-matrix defined as the average of the viability matrix over all the possible positions of s'' and l''. Mathematically we calculate the effective Y-matrix as following. We assume that for all the possible conformations of the 6-mer chain $s''s'sll'l''$ (see Fig. 10), there are N_i $(i=1,2,3,4,5,6)$ conformations of which the conformation of the 4-mer chain $s'sll'$ is of type-i defined in Fig. 7. For example, according to the 16 conformations shown in Fig. 10, N_1 is equal to 16. We also assume that for all these N_i conformations of $s''s'sll'l''$, there are n_{ij} conformations that can viably connect to a type-j conformation of the 4-mer chain $s'sll'$. We calculate the average Y-matrix using the following formula for the matrix element $y_{ij}(i,j=1,2,3,4,5,6)$:

$$y_{ij}=\frac{\Sigma_{m=1}^{N_i} n_{ij}+\Sigma_{m=1}^{N_j} n_{ji}}{N_i+N_j}\tilde{y}_{ij},$$

where $\tilde{y}_{ij}$ is defined in Eq. (3). In such a mean field approach, the matrix elements of the 6×6 matrix **Y** are in general noninteger numbers. The great advantage of this approach is that it keeps the computational efficiency while properly accounting for next-nearest-neighbor excluded volume.

We list the effective Y-matrix in the following for the connection between type-L,M,R,LR subunits (see Fig. 5). We use $Y_{t_1t_2}$ to denote the Y-matrix for the connection between a type-t_1 subunit and a type-t_2 subunit, where t_1 and t_2 are the subunit types defined in Fig. 5,

$$\mathbf{Y}_{tM}=\begin{bmatrix} 2 & 1.8 & 1.8 & 1.5 & 0.9 & 2.0 \\ 1.8 & 0.0 & 1.6 & 0.0 & 0.9 & 1.7 \\ 1.8 & 1.6 & 0.0 & 0.0 & 0.9 & 1.7 \\ 1.5 & 0.0 & 0.0 & 0.0 & 1.25 & 2.25 \\ 0.9 & 0.9 & 0.9 & 1.25 & 1.0 & 0 \\ 2.0 & 1.7 & 1.7 & 2.25 & 0 & 1.5 \end{bmatrix};$$

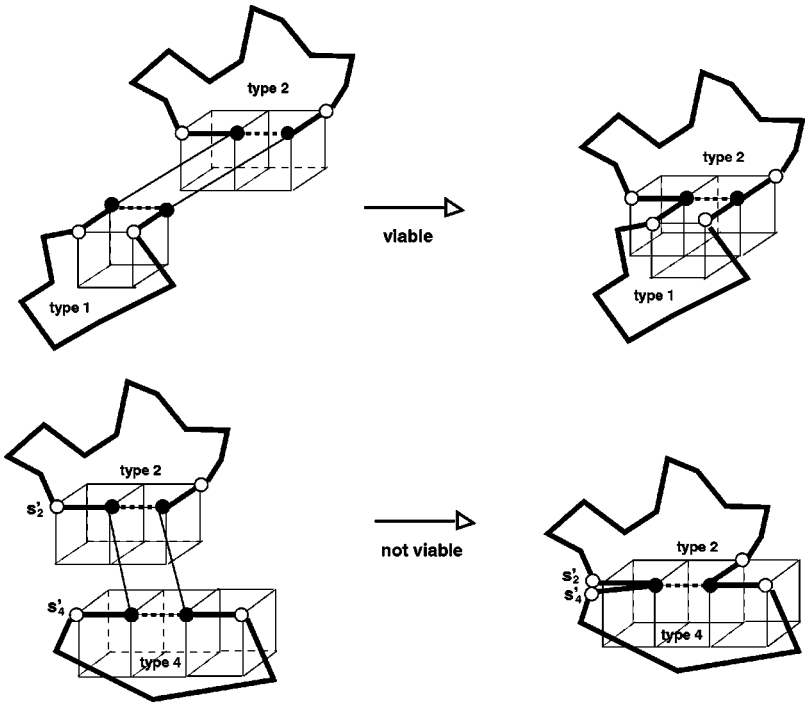

FIG. 8. The connection between the subunits. (A) The connection between type-2 and type-1 subunit conformations is viable; (B) the connection between type-2 and type-4 subunit conformations is not viable because the excluded volume of monomers s_2' and s_4'.

$$\mathbf{Y}_{tL}=\begin{bmatrix} 1.5 & 1.0 & 1.5 & 1.0 & 0.5 & 1.1 \\ 1.0 & 0.0 & 0.9 & 0.0 & 0.3 & 0.6 \\ 1.5 & 0.9 & 0.0 & 0.0 & 0.4 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.7 & 1.2 \\ 0.5 & 0.3 & 0.4 & 0.7 & 0.25 & 0.0 \\ 1.1 & 0.6 & 1.0 & 1.2 & 0.0 & 0.4 \end{bmatrix};$$

$$\mathbf{Y}_{tR}=\begin{bmatrix} 1.5 & 1.5 & 1.0 & 1.0 & 0.5 & 1.1 \\ 1.5 & 0.0 & 0.9 & 0.0 & 0.3 & 0.6 \\ 1.0 & 0.9 & 0.0 & 0.0 & 0.4 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.7 & 1.2 \\ 0.5 & 0.4 & 0.3 & 0.7 & 0.25 & 0.0 \\ 1.1 & 1.0 & 0.6 & 1.2 & 0.0 & 0.4 \end{bmatrix};$$

$$\mathbf{Y}_{LR}=\mathbf{Y}_{t,LR}=\begin{bmatrix} 1.0 & 0.5 & 0.6 & 0.7 & 0.3 & 0.6 \\ 0.5 & 0.0 & 0.33 & 0.0 & 0.0 & 0.2 \\ 0.6 & 0.33 & 0.0 & 0.0 & 0.1 & 0.2 \\ 0.7 & 0.0 & 0.0 & 0.0 & 0.2 & 0.5 \\ 0.3 & 0.0 & 0.1 & 0.2 & 0.14 & 0.0 \\ 0.6 & 0.2 & 0.2 & 0.5 & 0.0 & 0.0 \end{bmatrix}.$$

In the above formulas, t can be either type M or type I. Some of the other Y matrices can be obtained from the symmetry relation, $Y_{t_1t_2}=Y^T_{t_2t_1}$.

e. The count of conformations for a closed subgraph: Using the S and Y matrices, we can now derive an analytical formula for the calculation of the number of conformations $\Omega[x_1,y_1]$ for the closed subgraph $\{x_1,y_1\}$ in Fig. 1. The result $\Omega[x_1,y_1]$ can be conveniently calculated from a 6×6 G-matrix — $\mathbf{G_t}[\mathbf{x_1},\mathbf{y_1}]$ in the following equation:

$$\Omega[x_1,y_1]=\mathbf{U}_r\cdot\mathbf{G_t}[\mathbf{x_1},\mathbf{y_1}]\cdot\mathbf{U}_c; \tag{4}$$

where the G-matrix $\mathbf{G_t}[\mathbf{x_1},\mathbf{y_1}]$ is defined as a product of the S and Y matrices for the subunits,[16]

$$\mathbf{G_t}[\mathbf{x_1},\mathbf{y_1}]=\mathbf{S}_t^{(N)}\prod_{j=1}^{N-1}\mathbf{Y}\mathbf{S}^{(j)}, \tag{5}$$

where t ($t=L,M,R,\ldots,RRR$; see Fig. 5) is the type of the outermost subunit (shaded in Fig. 1), $\mathbf{S}_t^{(N)}$ is the S-matrix of the outermost subunit, N is the number of (nested) subunits in the graph, and $\mathbf{U}_r=\mathrm{row}[1,1,1,1,1,1]$; $\mathbf{U}_c=\mathrm{col}[1,1,1,1,1,1]$. $\mathbf{S}^{(j)}$ is the S-matrix of the jth subunit. For example, $\mathbf{S}^{(1)}$ is the S-matrix of the innermost subunit ($j=1$), and $\mathbf{S}_t^{(N)}$ is the S-matrix of the type-t outermost subunit ($j=N$).

The physical meaning of the G-matrix is the following: the row index i of the matrix element $G_t^{ij}[x_1,y_1]$ represents the type of the outlet conformation of the outermost subunit, and thus

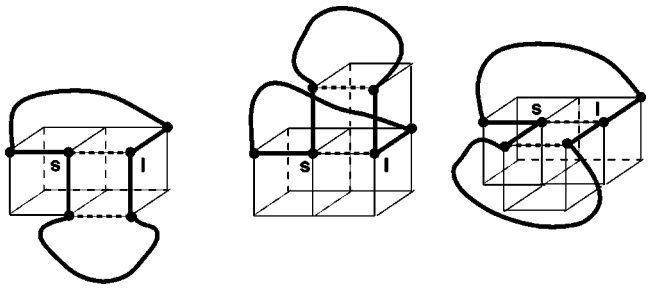

FIG. 9. There are multiple ways for the connection between two subunit conformations. This figure shows that there are three ways for the connection between a type-1 and a type-2 subunit conformation.

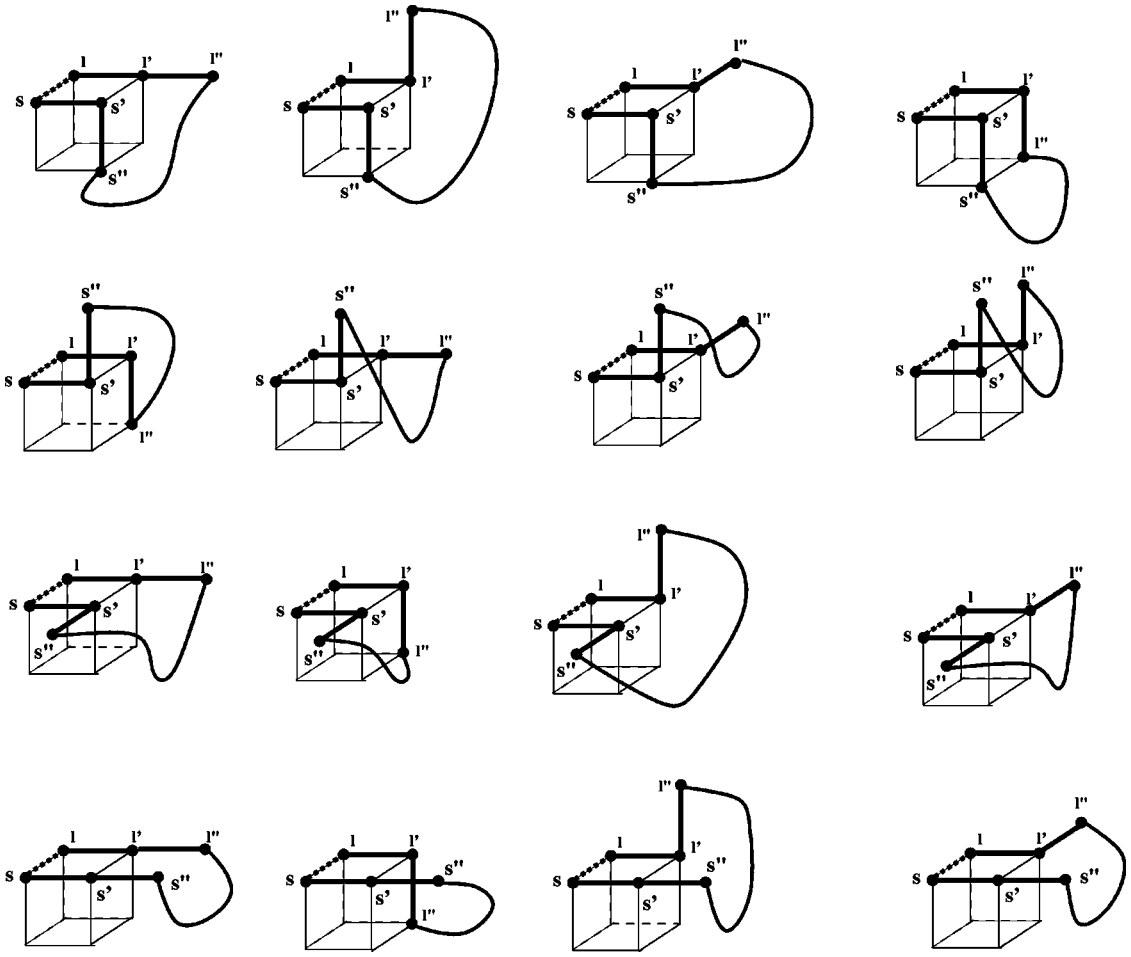

FIG. 10. The 16 conformations of a type-1 subunit conformation if we consider the excluded volume of the next-nearest-neighbor monomers s'' and l''.

$$g^i[x_1,y_1]=\sum_j G_t^{ij}[x_1,y_1] \tag{6}$$

is the number of conformations for the closed chain segment that has type-i conformation of the outlet (x_1,y_1).

B. The count of conformations for the whole graph

After obtaining the number of conformations for the closed subgraphs, we now consider the open subgraphs to complete the calculation for the whole polymer graph $\{x_0,y_0\}$ in Fig. 1. Note that the open parts (tails) $\{x_0,x_1\}$ and $\{y_1,y_0\}$ are connected to the closed chain segment $\{x_1,y_1\}$ at the outlet (x_1,y_1) of the outermost subunit. The connection is subject to the constraint that the open segments $\{x_0,x_1\}$ and $\{y_1,y_0\}$ cannot make contacts with the closed segment $\{x_1,y_1\}$ because there is no curved link between them on the graph. Therefore, the connection between the conformations of the closed and open segments (subgraphs) are effectively subject to strong exclusion interactions. We assume such exclusion interactions occur predominantly at the interface between the open and closed chain segments. For the polymer graph shown in Fig. 11, the interface is at the contact (x_1,y_1).

We assume that there are ω_i open chain conformations that are compatible to a type-i conformation ($i=1,2,\ldots,6$; see Fig. 7) of the contact (x_1,y_1). According to Eq. (6), we can write the count of conformations for the whole chain from monomer x_0 to y_0 as

$$\Omega=\sum_{i=1}^{6} \omega_i g^i[x_1,y_1]. \tag{7}$$

To calculate ω_i for different conformational type i, we classify the conformations of the tails into ''type-s'' (stiff) and ''type-f'' (flexible) according to the nearest and the next-nearest neighbor monomers near the interfacial monomers x_1 and y_1. A conformation is called type-s if the first two bonds form a straight line, and is called type-f otherwise.

We use $\omega_s(L)$ and $\omega_f(L)$ to denote the number of type-s and type-f conformations for a chain having L monomers. For $L\leqslant 16$, we obtain $\omega_s(L)$ and $\omega_f(L)$ by exhaustive enumeration on the cubic lattice (see Table II). For longer chains, $L>16$, we extrapolate the results from exact enumerations, and have

$$\omega_f(L)=-5.041+1.3997L+0.1277\log(L),$$

$$\omega_s(L)=-3.1629+1.4455L-0.483\log(L).$$

As shown in Table III, we can write ω_i as a function of $\omega_s(L_1),\omega_s(L_2),\omega_f(L_1)$, and $\omega_f(L_2)$, where $L1\equiv x_0-x_1+1$ and $L_2\equiv y_0-y_1+1$ are the lengths of the two tails. Us-

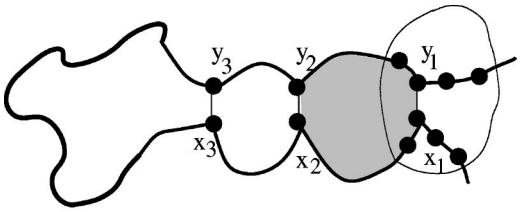

FIG. 11. Closed and open chain segments are connected at the outermost link (x_1,y_1) of the outermost subunit (shaded region); see Fig. 1. The viability of the connection is predominantly determined by the compatibility of the coupling of the conformations in the vicinity of the interface (x_1,y_1).

ing the above results for ω_i, from Eq. (7), we can obtain Ω, the number of chain conformations for the whole polymer graph in Fig. 1.

IV. THE PARTITION FUNCTION OF DOUBLE-STRANDED CHAIN CONFORMATIONS

A general polymer graph of double stranded chain conformations is composed of unrelated closed subgraphs [Fig. 12(A)], where each closed subgraph, embeds a set of closed sub-subgraphs [Fig. 12(B)]. A nested graph in Fig. 1 is a special case of the general polymer graph for double-stranded chain molecule in Fig. 12 with one embedded subgraphs ($N=1$) for the closed subgraphs.

The calculation of the partition function for the whole chain involves two steps: We first compute the partition function for the conformations of closed graphs $\{x,y\}$ in Fig. 12(A) by enumerating all the possible values of $x_1,y_1,x_2,y_2,\ldots,$ and x_N,y_N in Fig. 12(B), then enumerate possible positions of the unrelated closed subgraphs in Fig. 12(A) (e.g., x and y's) to obtain the partition function of the whole chain, including the partition functions of the open graphs.

A. The partition function of closed graphs

According to Eq. (4), the count of chain conformations for a closed graph is determined by the 6×6 G-matrix. Correspondingly, we introduce a 6×6 $\overline{G}$-matrix such that the partition function of the closed subgraph $\{x,y\}$ is given by

TABLE II. $\omega_s(L)$ and $\omega_f(L)$ for $L\leqslant16$.

L	$\omega_s(L)$	$\omega_f(L)$
1	1	0
2	1	0
3	1	1
4	5	3
5	21	11
6	89	41
7	369	161
8	1521	639
9	6277	2586
10	25785	10487
11	105901	42868
12	434341	174997
13	1778377	716960
14	7282557	2930694
15	29767953	12002201
16	122895565	49044152

TABLE III. ω_i $(i=1,2,3,4,5,6)$ as a function of $f_1\equiv\omega_f(L_1), f_2\equiv\omega_f(L_2), s_1\equiv\omega_s(L_1)$, and $s_2\equiv\omega_s(L_2)$, where $L1\equiv x_0-x_1+1$ and $L_2\equiv y_0-y_1+1$ are the lengths of the two tails in Fig. 1. For example, $(\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6)=(1,1,1,1,1,1)$ if $x=x_0$, $y=y_0$ and $\omega_3=(10f_1+4s_1)$ if $x\neq x_0$, $y=y_0$.

(x,x_0) (y_0,y)	$x=x_0$, $y=y_0$	$x=x_0$, $y\neq y_0$	$x\neq x_0$, $y=y_0$	$x\neq x_0$, $y\neq y_0$
ω_1	1	$10f_2+4s_2$	$10f_1+4s_1$	$30f_1f_2+12f_1s_2+12f_2s_1+6s_1s_2$
ω_2	1	$6f_2+3s_2$	$6f_1+3s_1$	$10f_1f_2+4f_1s_2+4f_2s_1+4s_1s_2$
ω_3	1	$10f_2+4s_2$	$10f_1+4s_1$	$10f_1f_2+4f_1s_2+4f_2s_1+4s_1s_2$
ω_4	1	$8f_2+4s_2$	$8f_1+4s_1$	$6f_1f_2+6f_1s_2+6f_2s_1+6s_1s_2$
ω_5	1	$7f_2+3s_2$	$7f_1+3s_1$	$4f_1f_2+3f_1s_2+3f_2s_1+s_1s_2$
ω_6	1	$9f_2+3s_2$	$9f_1+3s_1$	$11f_1f_2+6f_1s_2+6f_2s_1+3s_1s_2$

$$\mathbf{U}_r\cdot\overline{\mathbf{G}}_{\mathbf{t}}[\mathbf{x},\mathbf{y}]\cdot\mathbf{U}_c, \tag{8}$$

where $\mathbf{U}_r=\mathrm{row}[1,1,1,1,1,1]$; $\mathbf{U}_c=\mathrm{col}[1,1,1,1,1,1]$, and t $(t=L,M,R,\ldots,RRR)$ is the type of the subunit shaded in Fig. 12(B). Similar to the definition of the G-matrix, matrix element $\overline{G}_t^{ij}$ of a $\overline{G}$-matrix is defined as the (conditional) partition function for all the conformations for the closed subgraphs $\{x,y\}$, on the condition that the conformation of the outermost link (x,y) is of type-i and the conformation of the innermost link of subgraph $\{x_N,y_N\}$ is of type-j $(i,j=1,2,3,4,5,6)$.

Mathematically, the $\overline{G}$-matrix $\overline{\mathbf{G}}_{\mathbf{t}}[\mathbf{x},\mathbf{y}]$ is defined as the sum of the G-matrices weighted by the Boltzmann factor e^{-E/k_BT} (E = energy of the conformations corresponding to the graph) for all the possible closed subgraphs $\{x,y\}$ that have a type-t outermost subunit [shaded in Fig. 12(B)]. According to Eq. (5), $\overline{\mathbf{G}}_t[x,y]$ can be calculated as

$$\overline{\mathbf{G}}_t[x,y]=\sum_{\{x,y\}}\overline{\mathbf{S}}_t\cdot\mathbf{Y}\cdot\overline{\mathbf{K}}[x,y],$$

where $\Sigma_{\{x,y\}}$ denotes the sum over all the possible type-t closed graphs $\{x,y\}$, $\overline{\mathbf{S}}_t$ is the S-matrix of the shaded subunit in Fig. 12(B) weighted by the Boltzmann factor $e^{-\Delta E/k_BT}$ [ΔE = the energy change due to the formation of the contact (x,y)], and the $\overline{K}$-matrix $\overline{\mathbf{K}}[x,y]$ is defined as the product of the $\overline{G}$-matrices for the N embedded subgraphs,[17]

$$\overline{\mathbf{K}}[x,y]=\Sigma_{\{x_1,y_1\},\ldots,\{x_N,y_N\}}\overline{\mathbf{G}}_{t_1}[x_1,y_1]\cdot\prod_{i=2}^{N}\mathbf{Y}\cdot\overline{\mathbf{G}}_{t_i}[x_i,y_i],$$

where $\Sigma_{\{x_1,y_1\},\ldots,\{x_N,y_N\}}$ denotes the sum over all the possible values of x_i,y_i $(i=1,2,\ldots,N)$, and $\overline{\mathbf{G}}_{t_i}[x_i,y_i]$ is the $\overline{G}$-matrix of the subgraph $\{x_i,y_i\}$.

The $\overline{K}$-matrix $\overline{\mathbf{K}}_{\mathbf{t}}[\mathbf{x},\mathbf{y}]$ can be determined recursively from the $\overline{K}$-matrices for shorter chain segments. In Ref. 17, a detailed diagrammatic illustration was shown for 2D lattice chain conformations. For 3D lattice conformations, the basic idea is similar, but the results for the recursive relations are much more complex because 11 (for 3D) rather than 5 (for

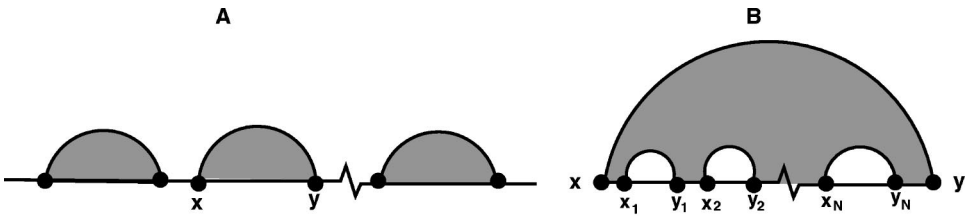

FIG. 12. (A) A general polymer graph for double-stranded chain conformations consists of unrelated closed subgraphs. (B) Each closed subgraph $[x,y]$ is composed of unrelated subgraphs: $[x_1,y_1],[x_2,y_2],\ldots,[x_N,y_N]$, embedded in the curved link (x,y).

2D) types of graphs are involved. We present the recursive relations for the $\overline{K}$-matrices for the 3D lattice conformations in the following equations:

$$\overline{\mathbf{K}}_L[x,y]=\overline{\mathbf{K}}_L[x,y-1]+\overline{\mathbf{K}}_{LR}[x,y-1]+\overline{\mathbf{K}}_{LRR}[x,y-1],$$

$$\overline{\mathbf{K}}_M[x,y]=\overline{\mathbf{K}}_M[x,y-1]+\overline{\mathbf{K}}_R[x,y-1]+\overline{\mathbf{K}}_{RR}[x,y-1]+\overline{\mathbf{K}}_{RRR}[x,y-1],$$

$$\overline{\mathbf{K}}_R[x,y]=\overline{\mathbf{K}}_R[x+1,y]+\overline{\mathbf{K}}_{LR}[x+1,y]+\overline{\mathbf{K}}_{LLR}[x+1,y],$$

$$\overline{\mathbf{K}}_{LL}[x,y]=\overline{\mathbf{K}}_{LL}[x,y-1]+\overline{\mathbf{K}}_{LLR}[x,y-1],$$

$$\overline{\mathbf{K}}_{RR}[x,y]=\overline{\mathbf{K}}_{RR}[x+1,y]+\overline{\mathbf{K}}_{LRR}[x+1,y],$$

$$\overline{\mathbf{K}}_{LR}[x,y]=\overline{\mathbf{G}}_M[x,y]+\overline{\mathbf{G}}_I[x,y] \quad (l=2),$$

$$\overline{\mathbf{K}}_{LR}[x,y]=\sum_{z=x+1}^{y-1}\{\overline{\mathbf{K}}_{LRR}[x,z]\cdot\mathbf{Y}\cdot\Sigma_{t_1=M,I,R,L}\overline{\mathbf{G}}_{t_1}[z,y]+\overline{\mathbf{K}}_{LR}[x,z]\cdot\mathbf{Y}\cdot\Sigma_{t_1=M,R,I}^{LR,L,LL}\overline{\mathbf{G}}_{t_1}[z,y]+\overline{\mathbf{K}}_L[x,z]\cdot\mathbf{Y}\cdot\Sigma_{t_1=M,I,R,LR}^{LLR,L,LL,LLL}\overline{\mathbf{G}}_{t_1}[z,y]\} \quad (l>2),$$

$$\overline{\mathbf{K}}_{LLR}[x,y]=\sum_{z=x+1}^{y-1}\{\overline{\mathbf{K}}_{LLR}[x,z]\cdot\mathbf{Y}\cdot\Sigma_{t_1=M,I,L}^{LL,R,LR}\overline{\mathbf{G}}_{t_1}[z,y]+\overline{\mathbf{K}}_{LL}[x,z]\cdot\mathbf{Y}\cdot\Sigma_{t_1=M,I,L,LL,LLL}^{R,LR,LLR}\overline{\mathbf{G}}_{t_1}[z,y]\},$$

$$\overline{\mathbf{K}}_{LRR}[x,y]=\sum_{z=x+1}^{y-1}\{\Sigma_{t_1=M,I,L}^{R,LR,RR}\overline{\mathbf{G}}_{t_1}[x,z]\cdot\mathbf{Y}\cdot\overline{\mathbf{K}}_{LRR}[z,y]+\Sigma_{t_1=M,I,L,R,LR}^{LRR,RR,RRR}\overline{\mathbf{G}}_{t_1}[x,z]\cdot\mathbf{Y}\cdot\overline{\mathbf{K}}_{RR}[z,y]\},$$

$$\overline{\mathbf{K}}_{LLL}[x,y]=\overline{\mathbf{K}}_{LLL}[x,y-1],$$

$$\overline{\mathbf{K}}_{RRR}[x,y]=\overline{\mathbf{K}}_{RRR}[x,y-1],$$

where t_1 in the above equations is the type of the closed graph $\{z,y\}$.

B. The partition function of the whole chain

To obtain the partition function for the conformations of the whole chain, we need to enumerate all the possible graphs shown in Fig. 12(A), including the open chain segments (i.e., chain segments that do not involve any contacts). The partition function of the whole chain from monomer x to y can be recursively calculated from the partition functions for shorter chains by growing the chain from x to y, with one monomer added for each step.

The viability of adding a monomer to the terminal of an existing chain conformation is mainly determined by the excluded volume near that terminal, which is effectively determined by the number of monomers already in contact with the terminal monomer. Therefore, we classify polymer graphs according to the number of curved links connected to the terminal monomer. We call a polymer graph to be of type-c if the terminal monomer has c contacting monomers (i.e., c curved links). In a 3D lattice model, viable c can be 0,1,2,3, and 4.

Similar to the $\overline{G}$-matrix defined in Eq. (8) for the partition function of the closed graphs $\{x,y\}$, we define a 6×6 Q-matrix, $\mathbf{Q}_c[x,y]$, for the partition function for all the possible type-c graphs, so the full partition function $Q(T)$ of the whole chain is given by

$$Q(T)=\sum_{c=0}^{4}\mathbf{U}_r\cdot\overline{\mathbf{Q}}_c[\mathbf{x},\mathbf{y}]\cdot\mathbf{U}_c. \tag{9}$$

A detailed diagrammatic illustration for the chain-growth recursive algorithm for $\mathbf{Q}_c[x,y]$ is described in Ref. 17 for 2D lattice conformations. Compared with the 2D model, the 3D model is much more complex because of the more convoluted polymer graphs involved. We present the recursive relations for Q-matrix in the following equations for $\mathbf{Q}_c[x,z]$ $(c=0,1,2,3,4; x<z<y)$:

$$\mathbf{Q}_0[x,z]=\alpha\cdot(\mathbf{Q}_0[x,z-1]+\mathbf{Q}_1[x,z-1]+\mathbf{Q}_2[x,z-1])+\mathbf{Q}_3[x,z-1],$$

where $\alpha=5.0$ to account for the rotational degrees of freedom of the covalent bond $[z-1,z]$;

$$\mathbf{Q}_1[x,z]=\sum_{t=L,M,LL,LLL}\overline{\mathbf{G}}_t[x,z]+\sum_{z_1=1}^{z-1}\sum_{c=0}^{3}\mathbf{Q}_c[x,z_1]\cdot\sum_{t'}\overline{\mathbf{G}}_{t'}[z_1,z],$$

where t is the type of graph $[x,z]$, t' is the type of subgraph $[z_1,z]$, and c is the type of graph $[x,z_1]$. $t'=LLL,LL,L,M$ for $c=0$, $t'=LL,L,M$ for $c=1$, $t'=L,M$ for $c=2$ and $t'=M$ for $c=3$;

$$\mathbf{Q}_2[x,z]=\sum_{t=R,LR,LLR}\overline{\mathbf{G}}_t[x,z]+\sum_{z_1=1}^{z-1}\sum_{c=0}^{2}\mathbf{Q}_c[x,z_1]\cdot\sum_{t'}\overline{\mathbf{G}}_{t'}[z_1,z],$$

where $t'=R,LR,LLR$ for $c=0$, $t'=R,LR$ for $c=1$ and $t'=R$ for $c=2$;

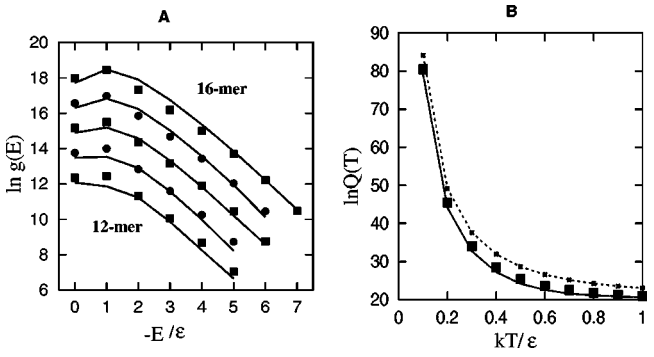

FIG. 13. (A) Test of the calculated (lines) and exactly enumerated (symbols) density of states $g(E)$ for 12-mers to 16-mers for double-stranded chain conformations on 3D lattice. Energy $E=$(number of contacts)$\cdot(-\epsilon)$. (B) Test of the theory (solid square) against exact enumeration (line) for the partition function $Q(T)$ of a 16-mer 3D lattice chain. The dashed line is the result predicted by the model with the Y-matrix accounting for only the nearest-neighbor monomers at the interface. Evidently, the dashed line cannot give reliable predictions.

$$\mathbf{Q}_3[x,z]=\sum_{t=RR,LRR}\overline{\mathbf{G}}_t[x,z]+\sum_{z_1=1}^{z-1}\sum_{c=0}^{1}\mathbf{Q}_c[x,z_1]\cdot\sum_{t'}\overline{\mathbf{G}}_{t'}[z_1,z],$$

where $t'=RR,LRR$ for $c=0$ and $t'=RR$ for $c=1$;

$$\mathbf{Q}_4[x,z]=\overline{\mathbf{G}}_{RRR}[x,z]+\sum_{z_1=1}^{z-1}\mathbf{Q}_0[x,z_1]\cdot\overline{\mathbf{G}}_{RRR}[z_1,z].$$

From Eq. (9), we obtain the partition function $Q(T)$ for the whole chain.

V. TESTS IN THE 3D LATTICE MODEL

We test the model by computing the partition functions for short chains, and compare the calculated results from the model with the results from the exact computer enumeration on the 3D lattice. We choose the simplest contact-based energy functions for the conformations; each contact contributes a sticking energy $-\epsilon$. In Fig. 13, we show the comparison between our predictions and the results from exact computer enumerations for a 16-mer homopolymer chain. We find that our model gives quite accurate results. As a comparison, in the figure, we also show the poor results from the Y-matrices that neglect the excluded volume of the nearest-neighbor monomers. From Fig. 13, we find that the excluded volume of the next-nearest-neighbor monomers plays an important role in determining the partition function.

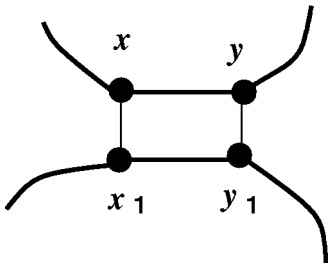

FIG. 14. A stack conformation formed by two consecutive contacts (x,x_1) and (y,y_1).

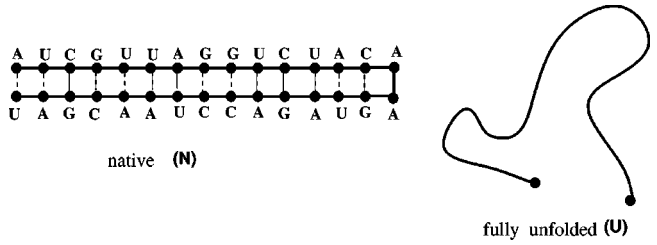

FIG. 15. The native structure (N) and completely unfolded structure (U) of the sequence AUCGUUAGGUCUACAAGUAGACCUAACGAU.

VI. 3D MODEL VS 2D MODEL FOR CONFORMATIONAL TRANSITIONS

A previous theory based on the 2D model has been applied to predict the thermodynamic properties of RNA secondary structures.[18] The 2D model gave the correct general profile of the thermal denaturation curves of RNA secondary structures, but over-estimated the breadth of the heat capacity melting curves, implying less sharp conformational transitions predicted by the model. One possible reason of the underestimation of the sharpness of the transition is the 2D representation of chain conformations. We now use the new 3D model to investigate the effect of dimensionality (3D vs 2D) of chain conformations on the thermodynamic properties of conformational transitions for double-stranded chains.

We first define the energy function. We assume that an RNA-like chain has a specific sequence of four types of monomers: A, U, C, and G, resembling the 4 types of bases of an RNA. We use the following base stacking energies: only stacks formed by the Watson–Crick (A–U and G–C) pairs contribute nonzero energies, and each A–U pair contributes an attractive energy $-\epsilon/2$, and each C–G pair contributes an attractive energy $-\epsilon$ to the stability of the stack. Figure 14 shows a stack formed by contacts (x,x_1) and (y,y_1).

As an illustrative calculation, we investigate the thermal melting transitions for a RNA sequence AUCGUUAGGUCUACAAGUAGACCUAACGAU. We choose this sequence

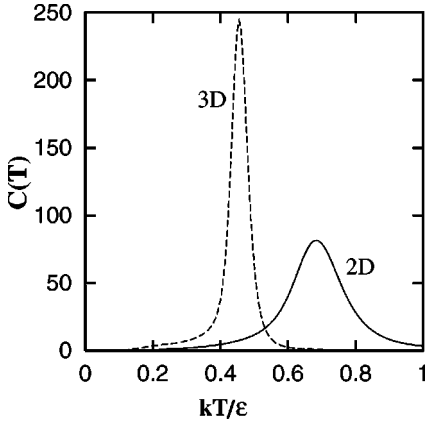

FIG. 16. The melting curve [heat capacity $C(T)$ vs temperature T] for sequence AUCGUUAGGUCUACAAGUAGACCUAACGAU.

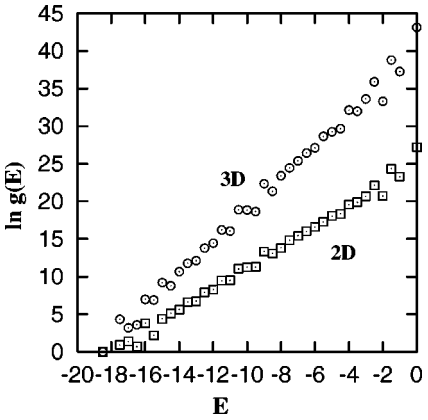

FIG. 17. The density of states for the sequence AUCGUUAGGUCUACAAGUAGACCUAACGAU.

because it has a unique native state (the lowest free energy state, see structure N in Fig. 15) and as we will see, the thermal melting transition is predominantly between two well defined states ($N \leftrightarrow U$ in Fig. 15).

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature T: $C(T) = (\partial/\partial T)[k_B T^2 (\partial/\partial T) \ln Q]$. The $C(T)$ vs T curves are shown in Fig. 16 for the 2D and 3D model. Both the 2D and 3D model predict a single peak melting curve. However, the melting temperature T_m predicted by the 3D model is lower than the T_m predicted by the 2D model, and the peak in the melting curve predicted by the 3D model is much sharper than that predicted by the 2D model, suggesting a more abrupt melting conformational transition in the 3D model. The notable sharpness of the melting curve predicted by the 3D model suggests that, as compared with the 2D model, the 3D model may give improved predictions for the RNA melting thermodynamics.

The differences in the thermodynamic properties between the 2D and 3D model arise from the fact that chain conformations in the 3D model have more degrees of freedom. To examine the distribution of the number of conformations over different energies, in Fig. 17, we show the results for the density of states $g(E)$ in 2D and 3D model respectively, where $g(E)$ is defined as the number of confor-

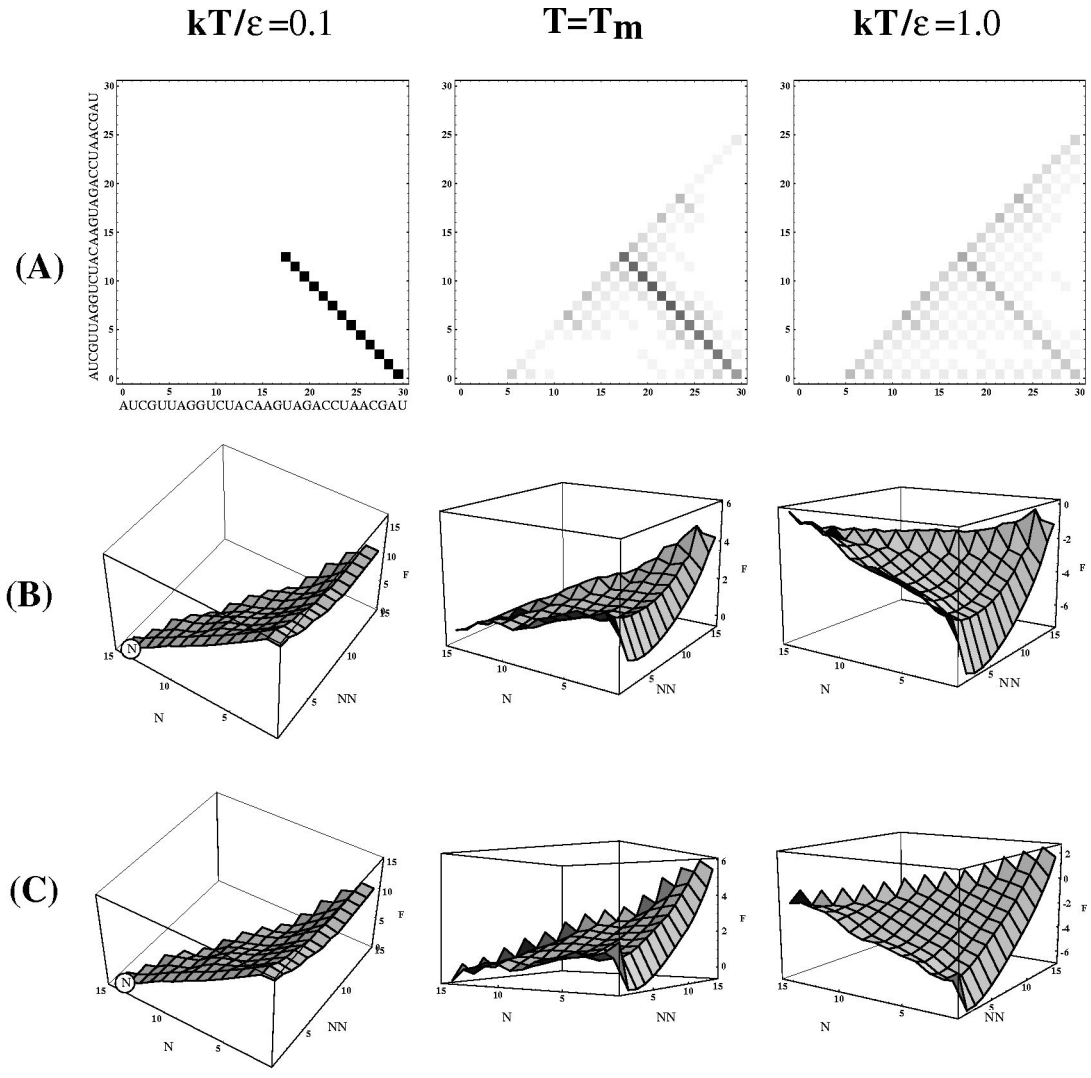

FIG. 18. (A) The density plots for the contact probability $p_{ij}(T)$ in the 3D model. The 2D model gives virtually the same density plots; (B) and (C) the free energy landscapes predicted by the 2D (B) and 3D (C) model.

mations that have energy E. We find that as the chain unfolds, the energy increases from $E_N=-18.5\epsilon$ for the native state N to $E_U=0$ for the completely unfolded state, the density of states increases from 1 to 5.33×10^{18} in the 3D model, and from 1 to 6.58×10^{11} in the 2D model. The overall increase in the density state is much steeper in the 3D model than in the 2D model. As a result, the melting temperature, crudely estimated by $k_BT_m\simeq[E_U-E_N]/[\ln g(E_U)/g(E_N)]$, is lower in the 3D model ($k_BT_m\simeq0.43\epsilon$) than in the 2D model ($k_BT_m\simeq0.68\epsilon$). We note that the estimated T_m's agree with the results from the melting curves quite accurately.

In order to gain a structural insight into the melting process, we calculate the contact probability $p(i,j)$ for every possible contact pair (i,j). In thermal equilibrium, $p_{ij}(T)$ is determined by the ratio of the conditional partition function $Q_{ij}(T)$ for all the conformations that contain contact (i,j) and the full partition function $Q(T)$ defined in Eq. (1), $p_{ij}(T)=Q_{ij}(T)/Q(T)$. From the contact probability $p_{ij}(T)$, we can derive the structure of the molecule at temperature T. In Fig. 18(A), we show the density plot for $p_{ij}(T)$. The 2D and 3D model virtually predicts the same conformational changes between the native state (N in Fig. 15) and the completely unfolded state (U). The $N\leftrightarrow U$ corresponds to the melting transition at the melting temperature T_m.

The 3D model predicts much sharper melting curves, does this suggest much higher cooperativity for the melting transition in the 3D model? We answer this question from the free energy landscape. A free energy landscape is defined as the free energy of the system as a function of the structures. To describe a structure mathematically, we define a contact in a conformation to be ''native'' if it also exists in the native state, and it is called ''non-native'' otherwise. We use parameters (N,NN) to describe the conformations, where N and NN are the number of native and non-native contacts of the conformation, respectively. The free energy landscape $F(N,NN,T)$ is calculated from the conditional partition function $Q(N,NN,T)$, defined as the partition function for all the conformations that have N native contacts and NN non-native contacts[18]: $F(N,NN,T)=-k_BT\ln Q(N,NN,T)$. The minima on the free energy landscape correspond to stable states. The less populous the intermediate states between the stable states, the more cooperative the conformational transition (between the stable states).

From the free energy landscapes shown in Figs. 18(B) and 18(C), we find that the barrier separating valley N and U at the melting temperature T_m is only slightly higher in the 3D model than in the 2D model, suggesting that the melting transition predicted by the 3D model is slightly more cooperative than the transition predicted by the 2D model.[21]

To quantify the cooperativity of the melting transition, we compute the van't Hoff enthalpy $\Delta H_{\rm vH}$ at the melting temperature T_m, and compare $\Delta H_{\rm vH}$ with the calorimetric enthalpy $\Delta H_{\rm cal}$ of the entire transition. Larger ratio $\kappa=\Delta H_{\rm vH}/\Delta H_{\rm cal}$ ($\kappa\leqslant1$) corresponds to higher cooperativity for conformational transition.[22–25]

The van't Hoff enthalpy $\Delta H_{\rm vH}$ can be calculated from the heat capacity $C(T)$ melting curve,[23–26] $\Delta H_{\rm vH}=2T_m\sqrt{k_BC(T_m)}$, which gives $\Delta H_{\rm vH}\simeq14.2\epsilon$ for the 3D model and $\Delta H_{\rm vH}\simeq12.3\epsilon$ for the 2D model. The calorimetric enthalpy $\Delta H_{\rm cal}$ of the entire melting process is given by $\Delta H_{\rm cal}=H(\infty)-H(0)$, where $H(T)=k_BT^2(d/dT)Q(T)$ is the enthalpy of the system at temperature T. We find that $\Delta H_{\rm cal}\simeq18.5\epsilon$ for both 3D and 2D model. Using the above $\Delta H_{\rm vH}$ and $\Delta H_{\rm cal}$, we find $\kappa\simeq0.77$ for the 3D model and $\kappa\simeq0.66$ for the 2D model. This result confirms our conclusion from the free energy landscapes that the melting transition predicted in the 3D model is slightly more cooperative than that predicted in the 2D model.

VII. SUMMARY

We have developed a 3D model for the statistical thermodynamics for the conformational changes of double-stranded chain molecules (e.g., RNA secondary structures). The theory relies on the polymer graph representation of the intrachain contacts. The partition function is computed as a product of matrices, where each matrix corresponds to the partition function of a subunit defined by the graph. The major differences between a 3D model and a 2D model stem from the count of the chain conformations. As compared with the 2D model, the 3D model has the following distinctive features: (1) Since each monomer can have more contacting neighbors in the 3D model than in the 2D model, there are more types of viable polymer graphs and conformational types for a given chain molecule, and hence the 3D model involves much more complex graphs and matrices for the partition functions; (2) Unlike the 2D model, the excluded volume of non-nearest-neighbor monomers at the interfaces of the subunits play important roles in determining the viability of the connection of subunit conformations. The new 3D model has been tested and validated against exact computer enumeration for lattice chain conformations.

Using the new 3D model, we investigate the effect of the dimensionality (3D vs 2D) of the chain representation on the sharpness and cooperativity of conformational transitions. An illustrative calculation shows that for a nearly two-state thermal melting transition, the 3D model gives much sharper melting curves and conformational transitions than the 2D model. Since the 2D model generally over-estimates the breadth of the melting curves for RNA secondary structures,[18] we expect that the new 3D model may give much improved predictions for the RNA melting thermodynamics.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the MU Research Board, The Petroleum Research Fund, and American Heart Association (National).

[1] Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
[2] D. Gront, A. Kolinski, and J. Skolnick, J. Chem. Phys. **113**, 5065 (2000).
[3] D. Thirumalai and D. K. Klimov, Folding Des. **3**, R112 (1998).
[4] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **96**, 6166 (1999).
[5] H. S. Chan and K. A. Dill, Annu. Rev. Biophys. Biophys. Chem. **20**, 447 (1991).
[6] H. S. Chan, S. Bromberg, and K. A. Dill, Philos. Trans. R. Soc. London, Ser. B **348**, 61 (1995).
[7] K. A. Dill *et al.*, Protein Sci. **4**, 561 (1995).

[8] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[9] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[10] M. Karplus and E. Shakhnovich, in *Protein Folding*, edited by T. E. Creighton (Freeman, New York, 1992), p. 127; D. Thirumalai, Theor. Chem. Acc. **103**, 292 (2000).
[11] W. L. Mattice and H. A. Scheraga, Biopolymers **23**, 1701 (1984).
[12] W. L. Mattice and U. W. Suter, *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems* (Wiley, New York, 1994).
[13] D. C. Poland and H. A. Scheraga, *Theory of the Helix–Coil Transition* (Academic, New York, 1970).
[14] J. Skolnick, Macromolecules **16**, 1069 (1982).
[15] J. Skolnick and E. Helfand, J. Chem. Phys. **72**, 5489 (1980).
[16] S.-J. Chen and K. A. Dill, J. Chem. Phys. **103**, 5802 (1995).
[17] S.-J. Chen and K. A. Dill, J. Chem. Phys. **109**, 4602 (1998).
[18] S.-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **97**, 646 (2000).
[19] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[20] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[21] J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[22] H. Kaya and H. S. Chan, Phys. Rev. Lett. **85**, 4823 (2000).
[23] H. Kaya and H. S. Chan, Proteins **40**, 637 (2000).
[24] H. S. Chan, Proteins **40**, 543 (2000).
[25] Y. Q. Zhou, C. K. Hall, and M. Karplus, Protein Sci. **8**, 1064 (1999).
[26] P. L. Privalov and S. A. Potekhin, Methods Enzymol. **131**, 4 (1986).