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

Wenbing Zhang and Shi-Jie Chen

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.

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 simulation approaches can accurately treat the atomic details of the chains, but often have the problems with conformational sampling, especially for large conformational transitions. 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. Polymer theory approach, e.g., the rotational isomeric states (RIS) models, 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 has been developed for a large class of chain conformations—double-stranded chain conformations, e.g., RNA secondary structures 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. 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, 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 \( Q(T) \),

\[
Q(T) = \sum_{\text{conf}} e^{-E/k_B T},
\]

where \( T \) is the temperature, \( k_B \) is the Boltzmann constant, \( \Sigma_{\text{conf}} \) 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 (\( \Sigma_{\text{conf}} \)).

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-
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 types 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.

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 $\beta$-hairpin structure. Since polymer graphs for general double-stranded chain conformations can be transformed into nested graphs through isomorphic transformations, the methodology developed here would also be applicable to general double-stranded conformations. 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

\[ Q(T) = \sum_{\text{graph}} \Omega \ e^{-E/k_B T}, \]

where $\Omega$ 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 $e^{-E/k_B T}$.

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 $\beta$ 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 [$\Omega$ 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 $\Omega$ 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 $\Omega$, the number of chain conformations for a given polymer graph.

\[ \text{FIG. 1. A nested polymer graph and the corresponding 3D cubic lattice chain conformation.} \]

\[ \text{Figs. 2, 3: Three types of relationships between two links of a polymer graph, and the corresponding chain conformations.} \]
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 (\( \Omega \) in Eq. (1)) involves two steps: we first count chain conformations for the closed subgraphs, then obtain \( \Omega \) 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 \( \Omega \), 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 \((x_1, y_3)\). 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, . . . , 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 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. 3. A schematic general double-stranded chain conformation and the corresponding polymer graph.
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 \((M \geq 18)\) subunit, the S-matrix is given by

\[
\ln s_{ij} = a_{ij}M + b_{ij} \ln M + c_{ij},
\]

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 \( M = 18 \). 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

\[
S \approx p(M) \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},
\]

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

d. The Y-matrix: Next we consider the nearest-neighbor approximation, the viability of the connection between subunit conformations is predominantly determined by the compatibility of the inlet-outlet combination. 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 \( s' \) and \( s'' \) 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-\( j \) and a type-\( i \) conformation.

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

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.
nearest-neighbors, we classify the subunit conformations into monomers, respectively. Therefore, by including the next-nearest-neighbor monomers, the subunit conformations can be classified into 20, 20, 25, 16, and 16 types. Similarly, we found that type-2, 3, 4, 5, and 6 subunit conformations of the next-nearest-neighbor monomers can be classified into 16 subtypes of conformations according to the Y-matrix. We use the table to calculate the average Y-matrix using the following formula for the matrix element $y_{ij}(i,j = 1,2,3,4,5,6)$:

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

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

We list the effective Y-matrix in the following for the connection between type-$M$ subunits because $a_{ij} = b_{ij} = c_{ij} = 0$ for $i = 1$ or $j = 1$.

<table>
<thead>
<tr>
<th>Type</th>
<th>ij</th>
<th>$a_{ij}$</th>
<th>$b_{ij}$</th>
<th>$c_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>2 2</td>
<td>1.40983</td>
<td>-1.99</td>
<td>-6.3452</td>
</tr>
<tr>
<td></td>
<td>3 3</td>
<td>1.40983</td>
<td>-1.99</td>
<td>-6.3452</td>
</tr>
<tr>
<td></td>
<td>4 4</td>
<td>1.122</td>
<td>3.46</td>
<td>-18.113</td>
</tr>
<tr>
<td></td>
<td>5 5</td>
<td>1.2078</td>
<td>1.9696</td>
<td>-14.3526</td>
</tr>
<tr>
<td></td>
<td>6 6</td>
<td>0.85</td>
<td>7.2</td>
<td>-22.31</td>
</tr>
<tr>
<td></td>
<td>7 7</td>
<td>1.47263</td>
<td>-3.15615</td>
<td>-4.37297</td>
</tr>
<tr>
<td></td>
<td>8 8</td>
<td>1.52125</td>
<td>-4.2039</td>
<td>-3.11496</td>
</tr>
<tr>
<td></td>
<td>9 9</td>
<td>0.9364</td>
<td>7.085</td>
<td>-25.6049</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.9364</td>
<td>7.085</td>
<td>-25.6049</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1.47404</td>
<td>-3.3118</td>
<td>-4.1796</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1.29522</td>
<td>0.122967</td>
<td>-11.8347</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.9364</td>
<td>7.085</td>
<td>-25.6049</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1.2131</td>
<td>1.9334</td>
<td>-15.8567</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.38873</td>
<td>1.5159</td>
<td>-8.2938</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1.47404</td>
<td>-3.3118</td>
<td>-4.1796</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>1.38873</td>
<td>-1.559</td>
<td>-8.2938</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>1.4231</td>
<td>-2.179</td>
<td>-6.55534</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>1.4231</td>
<td>-2.179</td>
<td>-6.55534</td>
</tr>
</tbody>
</table>

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 ($113 \times 113$) of the S and Y matrices. Therefore, we introduce an effective $6 \times 6$ 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 $s''s's'll'l''$ (see Fig. 10), there are $N_i$ $(i = 1,2,3,4,5,6)$ conformations of which the conformation of the 4-mer chain $s''s'll'$ is of type-$i$ defined in Fig. 7. For example, according to the 16 conformations shown in Fig. 10, $N_4$ is equal to 16. We also assume that for all these $N_i$ conformations of $s''s'll'l''$, there are $n_{ij}$ conformations that can viably connect to a type-$j$ conformation of the 4-mer chain $s''s'll'$. We calculate the average Y-matrix using the following formula for the matrix element $y_{ij}(i,j = 1,2,3,4,5,6)$:

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

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

$$Y_{LM} = \begin{bmatrix}
2 & 1.8 & 1.8 & 1.5 & 0.9 & 2.0 \\
1.8 & 0.0 & 1.6 & 0.0 & 0.9 & 1.7 \\
1.8 & 1.6 & 0.0 & 0.0 & 0.9 & 1.7 \\
1.5 & 0.0 & 0.0 & 0.0 & 1.25 & 2.25 \\
0.9 & 0.9 & 0.9 & 1.25 & 1.25 & 0 \\
2.0 & 1.7 & 1.7 & 2.25 & 0 & 1.5 
\end{bmatrix}.$$
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, \( Y_{tL}^T = Y_{tR} \).

\[ Y_{tL} = \begin{bmatrix} 1.5 & 1.0 & 1.5 & 1.0 & 0.5 & 1.1 \\ 1.0 & 0.0 & 0.9 & 0.0 & 0.3 & 0.6 \\ 1.5 & 0.9 & 0.0 & 0.0 & 0.4 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.7 & 1.2 \\ 0.5 & 0.3 & 0.4 & 0.7 & 0.25 & 0.0 \\ 1.1 & 0.6 & 1.0 & 1.2 & 0.0 & 0.4 \end{bmatrix} \]

\[ Y_{tR} = \begin{bmatrix} 1.5 & 1.0 & 1.5 & 1.0 & 0.5 & 1.1 \\ 1.5 & 0.0 & 0.9 & 0.0 & 0.3 & 0.6 \\ 1.0 & 0.9 & 0.0 & 0.0 & 0.4 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.7 & 1.2 \\ 0.5 & 0.4 & 0.3 & 0.7 & 0.25 & 0.0 \\ 1.1 & 1.0 & 0.6 & 1.2 & 0.0 & 0.4 \end{bmatrix} \]

\[ Y_{LR} = Y_{tL}^T Y_{tR} = \begin{bmatrix} 1.0 & 0.5 & 0.6 & 0.7 & 0.3 & 0.6 \\ 0.5 & 0.0 & 0.33 & 0.0 & 0.0 & 0.2 \\ 0.6 & 0.33 & 0.0 & 0.0 & 0.1 & 0.2 \\ 0.7 & 0.0 & 0.0 & 0.0 & 0.2 & 0.5 \\ 0.3 & 0.0 & 0.1 & 0.2 & 0.14 & 0.0 \\ 0.6 & 0.2 & 0.2 & 0.5 & 0.0 & 0.0 \end{bmatrix} \]

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, \( Y_{t'_{t''}}^T = Y_{t''_{t'}} \).

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 \( \times \) 6 G-matrix — \( G_{[x_1,y_1]} \) in the following equation:

\[ \Omega[x_1,y_1] = U_r \cdot G_{[x_1,y_1]} \cdot U_c; \]

where the G-matrix \( G_{[x_1,y_1]} \) is defined as a product of the \( S \) and \( Y \) matrices for the subunits,

\[ G_{[x_1,y_1]} = \prod_{j=1}^{N-1} Y S^{(j)}, \]

where \( t \) (\( t = L, M, R \), \ldots, \( RRR \); see Fig. 5) is the type of the outermost subunit (shaded in Fig. 1), \( S_{t}^{(N)} \) is the S-matrix of the outermost subunit, \( N \) is the number of (nested) subunits in the graph, and \( U_r = \text{row}[1,1,1,1,1,1]; \quad U_c = \text{col}[1,1,1,1,1,1]. \) \( S^{(j)} \) is the S-matrix of the \( j \)-th subunit. For example, \( S^{(1)} \) is the S-matrix of the innermost subunit (\( j = 1 \)), and \( S_{t}^{(N)} \) 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 \( G_{xy}^{[x_1,y_1]} \) represents the type of the outlet conformation of the outermost subunit, and thus
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 \([y_0, y_1]\) 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 \([y_0, y_1]\) 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 \(\omega_i\) 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

\[
\Omega = \sum_{i=1}^{6} \omega_i g_i'[x_1, y_1].
\]

To calculate \(\omega_i\) 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 \(L \leq 16\), 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 \(\omega_i\) as a function of \(\omega_s(L_1), \omega_f(L_2), \omega_s(L_1),\) and \(\omega_f(L_2),\) where \(L_1 = x_0 - x_1 + 1\) and \(L_2 = y_0 - y_1 + 1\) are the lengths of the two tails. Us-
TABLE II. \( \omega_i(L) \) and \( \omega_j(L) \) for \( L \leq 16 \).

<table>
<thead>
<tr>
<th>( L )</th>
<th>( \omega_i(L) )</th>
<th>( \omega_j(L) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>89</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>369</td>
<td>161</td>
</tr>
<tr>
<td>8</td>
<td>1521</td>
<td>639</td>
</tr>
<tr>
<td>9</td>
<td>6277</td>
<td>2586</td>
</tr>
<tr>
<td>10</td>
<td>25785</td>
<td>10487</td>
</tr>
<tr>
<td>11</td>
<td>105901</td>
<td>42868</td>
</tr>
<tr>
<td>12</td>
<td>434341</td>
<td>174997</td>
</tr>
<tr>
<td>13</td>
<td>1778377</td>
<td>716960</td>
</tr>
<tr>
<td>14</td>
<td>7282557</td>
<td>2930694</td>
</tr>
<tr>
<td>15</td>
<td>29767953</td>
<td>12002201</td>
</tr>
<tr>
<td>16</td>
<td>122895565</td>
<td>49044152</td>
</tr>
</tbody>
</table>

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 visibility of the connection is predominantly determined by the compatibility of the coupling of the conformations in the vicinity of the interface \((x_1,y_1)\).

IV. THE PARTITION FUNCTION OF DOUBLE-STRAINED 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 \(x_1, y_1, x_2, y_2, \ldots, x_N, y_N\) 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 \(6 \times 6\) \(G\)-matrix. Correspondingly, we introduce a \(6 \times 6\) \(G\)-matrix such that the partition function of the closed subgraph \(\{x, y\}\) is given by

\[
\mathbf{G}[x,y] = \sum_{\{s_i\}} \mathbf{S}_i \cdot \mathbf{Y} \cdot \mathbf{K}[x,y],
\]

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

\[
\mathbf{K}[x,y] = \mathbf{S}_1 \cdot \mathbf{G}_{i_1}[x_1,y_1] \cdot \mathbf{G}_{i_2}[x_2,y_2] \cdot \cdots \cdot \mathbf{G}_{i_N}[x_N,y_N]
\]

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

The \(K\)-matrix \(\mathbf{K}[x,y]\) can be determined recursively from the \(K\)-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

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

\[
\begin{align*}
\tilde{K}_L[x,y] &= \tilde{K}_L[x,y-1] + \tilde{K}_{LR}[x,y-1] + \tilde{K}_{LRR}[x,y-1], \\
\tilde{K}_M[x,y] &= \tilde{K}_M[x,y-1] + \tilde{K}_R[x,y-1] + \tilde{K}_{RR}[x,y-1] \\
&\quad + \tilde{K}_{RRR}[x,y-1], \\
\tilde{K}_R[x,y] &= \tilde{K}_R[x+1,y] + \tilde{K}_{LR}[x+1,y] + \tilde{K}_{LRR}[x+1,y], \\
\tilde{K}_{LL}[x,y] &= \tilde{K}_{LL}[x,y-1] + \tilde{K}_{LRR}[x,y-1], \\
\tilde{K}_{RR}[x,y] &= \tilde{K}_{RR}[x+1,y] + \tilde{K}_{LRR}[x+1,y], \\
\tilde{G}_M[x,y] &= \tilde{G}_M[x,y] + \tilde{G}_R[x,y] \quad (l=2), \\
\tilde{G}_{LR}[x,y] &= \sum_{z=x+1}^{y-1} \{ \tilde{G}_{LLR}[x,z] \cdot Y \cdot \tilde{G}_{R}[z,y] \\
&\quad + \tilde{K}_{L}[x,z] \cdot Y \cdot \sum_{l=1}^{3} \tilde{G}_{L}[z,y] \cdot \tilde{G}_{L}[z,y] \} \quad (l \geq 2), \\
\tilde{K}_{LLR}[x,y] &= \sum_{z=x+1}^{y-1} \{ \tilde{K}_{L}[x,z] \cdot Y \cdot \sum_{l=1}^{3} \tilde{G}_{L}[z,y] \cdot \tilde{G}_{L}[z,y] \\
&\quad + \sum_{l=1}^{3} \tilde{K}_{L}[z,y] \cdot Y \cdot \sum_{l=1}^{3} \tilde{G}_{L}[z,y] \cdot \tilde{G}_{L}[z,y] \}, \\
\tilde{K}_{LRR}[x,y] &= \sum_{z=x+1}^{y-1} \{ \tilde{K}_{L}[x,z] \cdot Y \cdot \sum_{l=1}^{3} \tilde{G}_{L}[z,y] \cdot \tilde{G}_{L}[z,y] \\
&\quad + \sum_{l=1}^{3} \tilde{K}_{L}[z,y] \cdot Y \cdot \sum_{l=1}^{3} \tilde{G}_{L}[z,y] \cdot \tilde{G}_{L}[z,y] \}, \\
\tilde{K}_{LLL}[x,y] &= \tilde{K}_{LLL}[x,y-1], \\
\tilde{K}_{RRR}[x,y] &= \tilde{K}_{RRR}[x,y-1],
\end{align*}
\]

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

**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 $G$-matrix defined in Eq. (8) for the partition function of the closed graphs $\{x, y\}$, we define a $6 \times 6$ Q-matrix, $\tilde{Q}_c[x, y]$, 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

\[
Q(T) = \sum_{c=0}^{4} U_c \cdot \tilde{Q}_c[x, y] \cdot U_c.
\]

A detailed diagrammatic illustration for the chain-growth recursive algorithm for $\tilde{Q}_c[x, y]$ 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 $\tilde{Q}_c[x, y] \quad (c=0, 1, 2, 3, 4; x < y < y)$:

\[
\begin{align*}
\tilde{Q}_0[x, z] &= \alpha \cdot (\tilde{Q}_0[x, z-1] + \tilde{Q}_1[x, z-1] + \tilde{Q}_2[x, z-1]), \\
\tilde{Q}_3[x, z] &= \sum_{t=L, M, L, L, L} \tilde{G}_{t}[x, z] + \sum_{z_1=1}^{3} \sum_{c=0}^{3} \tilde{Q}_c[x, z_1] \\
&\quad \cdot \sum_{t'} \tilde{G}_{t'}[z_1, z],
\end{align*}
\]

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 $[x, z_1]$. $t' = LLL, L, L, M$ for $c = 0$, $t' = LL, L, M$ for $c = 1$, $t' = L, M$ for $c = 2$ and $t' = M$ for $c = 3$;

\[
\tilde{Q}_2[x, z] = \sum_{t=R, L, R, L, L, L} \tilde{G}_{t}[x, z] + \sum_{z_1=1}^{2} \sum_{c=0}^{3} \tilde{Q}_c[x, z_1] \\
&\quad \cdot \sum_{t'} \tilde{G}_{t'}[z_1, z],
\]

where $t' = R, L, R, L, L, L, L$ for $c = 0$, $t' = R, L, R$ for $c = 1$ and $t' = R$ for $c = 2$.
nearest-neighbor monomers. From Fig. 13, we find that the
Y-matrices that neglect the excluded volume of the
comparison, in the figure, we also show the poor results from
We find that our model gives quite accurate results. As a
computer enumerations for a 16-mer homopolymer chain.
son between our predictions and the results from exact
energy functions for the conformations; each contact contrib-
on the 3D lattice. We choose the simplest contact-based en-
for short chains, and compare the calculated results from the
model with the Y-matrix accounting for only the
nearest-neighbor monomers at the interface. Evidently, the dashed line can-
result predicted by the model with the Y-matrix accounting for only the

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 \( E = \text{(number of contacts)} \times (-\epsilon) \).
(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.

\[
Q_3(x,z) = \sum_{t=RR, LRR} \bar{G}_t(x,z) + \sum_{c=0}^{z-1} \sum_{z_1=1}^1 Q_c(x,z_1)
\]

where \( t' = RR, LRR \) for \( c = 0 \) and \( t' = RR \) for \( c = 1 \);

\[
Q_4(x,z) = \bar{G}_{RRR}[x,z] + \sum_{z_1=1}^{z-1} Q_0(x,z_1) \cdot \bar{G}_{RRR}[z_1,z].
\]

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 contrib-
utes an attractive energy \( -\epsilon \). In Fig. 13, we show the compar-
ison 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.

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

A previous theory based on the 2D model has been ap-
plied to predict the thermodynamic properties of RNA sec-
ondary 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 consequentiol trans-
itions 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 con-
tributes an attractive energy \( -\epsilon/2 \), and each C–G pair con-
tributes 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 \text{AUCGUUAGGUCUACAAGUAGACCUAACGAU}. We choose this sequence

\[
\bar{G}_{RRR}[x,z] = \bar{G}_{RRR}[x,z] + \sum_{z_1=1}^{z-1} Q_0(x,z_1) \cdot \bar{G}_{RRR}[z_1,z]
\]

\[
Q_4(x,z) = \bar{G}_{RRR}[x,z] + \sum_{z_1=1}^{z-1} Q_0(x,z_1) \cdot \bar{G}_{RRR}[z_1,z].
\]

\[
Q_3(x,z) = \sum_{t=RR, LRR} \bar{G}_t(x,z) + \sum_{c=0}^{z-1} \sum_{z_1=1}^1 Q_c(x,z_1)
\]

\[
Q_4(x,z) = \bar{G}_{RRR}[x,z] + \sum_{z_1=1}^{z-1} Q_0(x,z_1) \cdot \bar{G}_{RRR}[z_1,z].
\]
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→U in Fig. 15).

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-

We first apply the 2D and 3D model to calculate the heat capacity $C(T)$ as a function of temperature $T$: $C(T) = (\partial / \partial T) \left[ k_B T^2 (\partial / \partial T) \ln Q \right]$. 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-
motions 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 5.33x10$^{18}$ in the 3D model, and from 1 to 6.58x10$^{11}$ 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 $k_B T_m = [E_U - E_N]/[\ln g(E_U)/g(E_N)]$, is lower in the 3D model ($k_B T_m = 0.43 \epsilon$) than in the 2D model ($k_B T_m = 0.68 \epsilon$). 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→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$^{18}$: $F(N,NN,T) = -k_B T \ln Q(N,NN,T)$. 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 $\Delta H_{vH}$ at the melting temperature $T_m$, and compare $\Delta H_{vH}$ with the calorimetric enthalpy $\Delta H_{cal}$ of the entire transition. Larger ratio $\kappa = \Delta H_{vH}/\Delta H_{cal}$ ($\kappa \approx 1$) corresponds to higher cooperativity for conformational transition.$^{22-25}$

The van't Hoff enthalpy $\Delta H_{vH}$ can be calculated from the heat capacity $C(T)$ melting curve,$^{23-26}$

$$\Delta H_{vH} = 2T_m^2 k_B C(T_m),$$

which gives $\Delta H_{vH} = 14.2 \epsilon$ for the 3D model and $\Delta H_{vH} = 12.3 \epsilon$ for the 2D model. The calorimetric enthalpy $\Delta H_{cal}$ of the entire melting process is given by $\Delta H_{cal} = H\left(\infty\right) - H(0)$, where $H(T) = k_B T^2 (dH/dT) Q(T)$ is the enthalpy of the system at temperature $T$. We find that $\Delta H_{cal} = 18.5 \epsilon$ for both 3D and 2D model. Using the above $\Delta H_{vH}$ and $\Delta H_{cal}$, we find $\kappa = 0.77$ for the 3D model and $\kappa = 0.66$ 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).