Predicting free energy landscapes for complexes of 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 19 September 2000; accepted 11 December 2000)

We develop a statistical mechanical theory for the free energy landscapes for complexes of double-stranded chain molecules. The theory is based on the generalized polymer graph, a graphical representation for the conformations of the complexes. We compute the partition functions by ''dividing and conquering'' on the generalized polymer graph: we decompose a graph into simple subunits, calculate the partition function of each subunit exactly, and treat the interactions between subunits approximately, by calculating the localized interactions (of the nearest neighbor and the next-nearest neighbor monomers) at the interface of subunits. Our tests against the exact computer enumeration on the two-dimensional (2D) square lattice show that the theory is accurate. We apply the theory to the computation of the free energy landscapes of three representative systems: homopolymer–homopolymer, homopolymer–heteropolymer, and heteropolymer–heteropolymer complexes, using contact-based energy functions for the homopolymer–homopolymer and homopolymer–heteropolymer complexes, and stacking energies for the heteropolymer–heteropolymer complexes (to mimic RNA secondary structures). We find that the systems involving homopolymers show smooth free energy landscapes, and undergo noncooperative structural transitions during the melting process, and that the system of heteropolymers show rugged free energy landscapes, and the thermal denaturation involves intermediate states and cooperative structural transitions. We believe this approach maybe useful for computing the free energy landscapes and the thermodynamics of DNA or RNA interactions and RNA binding to a DNA or RNA target. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1345722]

I. INTRODUCTION

Cellular functions at the molecular level involve binding of biomolecules. For example, enzymes and substrates from functional complexes to catalyze biochemical reactions, RNA molecules bind partially or completely in gene expression,[1] splicing,[2] and antisense regulation,[3] and DNA strands associate and dissociate in the duplication of DNA. Moreover, the rapidly developing *in vitro* DNA–RNA or DNA–DNA hybridization microarray technology critically relies on the formation of the bound complexes of molecules.[4]

How to predict the binding affinity, stability, stable native, and intermediate states of the bound complex? We treat biomolecules as heteropolymers. Current approaches to the binding process often assumes that the molecules have only one or few ''optimal'' or ''suboptimal'' structures, and the bound complex is formed through simple docking of these structures.[5,6] As a result of such approximations, the molecular complex has only a single optimal structure or few suboptimal structures. Such assumptions cause errors[7] because (1) the thermodynamic properties of the binding process is determined not only by few optimal and suboptimal structures, but also by the fluctuations of these structures, and by the whole ensemble of intermolecular and intramolecular structures, and (2) the structures of the complex are often formed through ''induced fit''—structures of the molecules undergo rearrangement upon binding.[8,9,10] Therefore, to reliably predict the thermodynamic properties of the molecular complexes, we must consider the full ensemble of conformations for the complex of molecules.

Brute-force atomic modes, like molecular dynamics or Monte Carlo simulations, are limited by the incomplete sampling of the conformational space. On the other hand, the ''simplified models,'' for example, the lattice mode,[11] can capture the full conformational ensemble for short chains, but they are limited to short chains with simplified chain representations, and cannot give quantitative predictions for realistic biopolymers.

Recently, a third type of model has been developed. The new model is based on the graph theoretic approach to chain conformational representation.[12,13,14] It treats conformational ensemble in an analytical and rigorous way. The model has been tested and validated both by the two-dimensional (2D) lattice models[13] and by the experimental thermodynamic measurements.[14] However, the applicability of the model is limited to intramolecular folding, it is unable to treat the folding and unfolding of intermolecular complexes. In the present paper, we develop a new model to predict the statistical thermodynamics of molecular complexes.

Models have been developed to calculate the dissociation thermodynamics of DNA double strands, based on the

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

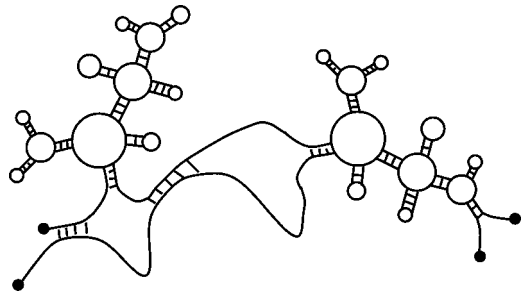

FIG. 1. A bound complex of two double stranded chain molecules.

Zimm–Bragg[15,16] type of helix–coil transition theory.[16–19] These models have been successful in predicting melting curves for short DNA duplexes under certain experimental conditions. However their predictions become less successful for long chains.[17,18] One possible reason is that the helix–coil transition theory do not treat complex chain conformations with multiple branched nonlocal intrachain and interchain contacts that occur for long chains, such as the structure shown in Fig. 1. Moreover, the helix–coil transition theory cannot accurately account for the sequence-dependent chain conformational entropy, because, for example, it neglects the sequence dependence of the loop entropy.[20,21]

In this paper, we describe the first steps toward a general statistical mechanical theory of the free energy landscape for the complexes of biopolymers. We define a bound complex of chain molecules as a system of two molecules that contain at least one intermolecular contact. Our theory is based on the recently developed graph theoretic approach to the calculation of the partition function of chain molecules.[12,13] From the partition functions, we compute the free energy landscapes, from which we can predict the stability and the structural changes for the melting of the complex prior to dissociation.[14,22,23,24] Because the general methodology of the graphic approach is independent of any specific chain representation, the present theory is applicable to chain molecules with any given representations for conformations and energy functions, although we provide only lattice model test in this paper.

II. FREE ENERGY LANDSCAPES OF BOUND COMPLEXES

The association and dissociation of two molecules, as shown in Fig. 2, can be described by a ''chemical equilibrium,''

$$S_1 + S_1' \rightleftharpoons S_2,$$

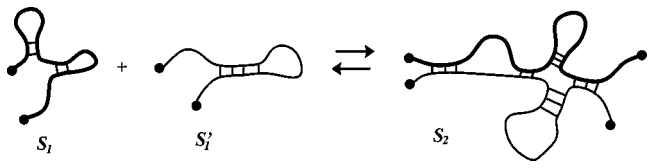

FIG. 2. Formation and dissociation of a bound complex S_2 consisting of two chain molecules S_1 and S_1'.

where ''reactants'' S_1 and S_1' are the (dissociated) free molecules, and the ''product'' S_2 are the associated molecular complex. The thermodynamic system consists of chain molecules S_1, S_1', and S_2, each having an ensemble of the chain conformations.

The stability and stable states of a thermodynamic system are determined by the free energy of the system. A powerful mathematical quantity that directly relates the free energy to the structures is the free energy landscape.[23,24,25] It is defined as the free energy $F(\mathbf{x})$ for all the possible structures described by structural parameters $\mathbf{x}$. $\mathbf{x}$ can be, for example, the number of intrachain and interchain contacts for a chain molecules.

A free energy landscape gives the distribution of all the structures: stable states correspond to the valleys, and unstable states correspond to the hills of the landscape. The native structure corresponds to the global minimum (the deepest valley) on the landscape, an the intermediate metastable structures correspond to the local minima (valleys). The free energy landscape depends on external parameters, such as temperature and solution condition. Changing the external parameters will lead to the change of the shape of the free energy landscape, and the change of the stable and metastable states. Therefore, the free energy landscape is the single mathematical quantity that describes all the conformations and their stabilities of a biomolecule.

The free energy $F(\mathbf{x})$ comes from the partition function $Z(\mathbf{x})$: $F(\mathbf{x}) = -kT \ln Z(\mathbf{x})$, $Z(\mathbf{x})$ is the weighted sum over all the possible states,

$$Z(\mathbf{x}) = \Sigma_{\text{conf}\,\mathbf{x}} e^{-E/kT},$$

where k is the Boltzmann constant, T is the temperature, Σ is the sum over all the possible states described by $\mathbf{x}$, including translational, rotational, and conformational degrees of freedoms, and E is the energy of an individual state.

Because we focus on the stability and conformational changes of the complex, we define the free energy landscapes $\Delta F(\mathbf{x})$ in terms of the structural parameter $\mathbf{x}$ of the complex, and choose the state of the free molecules as the reference state,

$$\Delta F(\mathbf{x}) = F_2(\mathbf{x}) - F_1 - F_1' = -kT \ln\left\{\frac{Z_2(\mathbf{x})}{Z_1 Z_1'}\right\}, \tag{1}$$

where F_1, F_1', and $F_2(\mathbf{x})$ are the free energies of the free molecules S_1, S_1' and the complex S_2, respectively, and Z_1, Z_1', and $Z_2(\mathbf{x})$ are the corresponding partition functions. The formation of a complex structure described by parameter $\mathbf{x}$ is favorable than the separated state if $\Delta F(\mathbf{x}) < 0$, and is unfavorable otherwise.

If we neglect the correlation between the chain conformation and the translational and rotational degrees of freedom, we can write the chain partition function as[16,18]

$$Z_2(\mathbf{x}) = Z_2^{(\text{ext})} Q_2(\mathbf{x}); \quad Z_1 = Z_1^{(\text{ext})} Q_1; \quad Z_1' = Z_1'^{(\text{ext})} Q_1',$$

where $Z_2^{(\text{ext})}$, $Z_1^{(\text{ext})}$, and $Z_1'^{(\text{ext})}$ are the translational and rotational partition functions, and $Q_2(\mathbf{x})$, Q_1, and Q_1' are the chain conformational partition functions. Letting

$$\alpha = Z_1^{(\text{ext})} Z_1'^{(\text{ext})} / Z_2^{(\text{ext})} \tag{2}$$

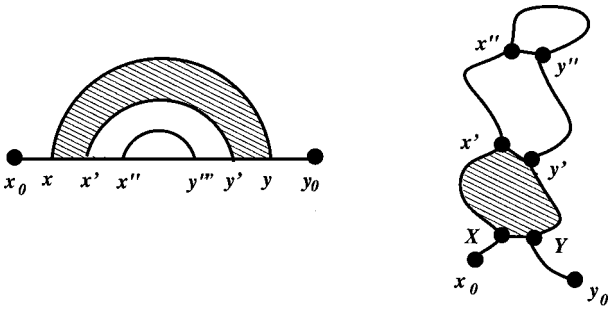

FIG. 3. A polymer graph (left) and the corresponding chain conformation (right). The shaded region is a subunit in the polymer graph.

leaves

$$\Delta F(\mathbf{x}) = -kT \ln\left\{\frac{Q_2(\mathbf{x})}{\alpha Q_1 Q_1'}\right\}. \tag{3}$$

α is related to the nucleation free energy for the formation of a complex,[18,26] it contains the concentration-dependence and other factors associated to the translational and rotational degrees of freedom, for example, the dependence of the rotational partition function on the principle moments of inertia of the molecules.[16,17,18] As a first order approximation, we neglect the temperature dependence of α.

To obtain the free energy landscape $\Delta F(\mathbf{x})$, we need to calculate the partition functions $Q_2(\mathbf{x})$ of the complex and Q_1 and Q_1' of the free molecule. A graph theoretic approach has been previously developed to compute the partition functions for free molecules.[12,13,14] In this work, we develop a theory to compute the partition function of the complex.

III. THE PARTITION FUNCTION OF FREE MOLECULES

We first briefly review the graph theoretic approach to the partition functions of the free molecules. Details of the theory can be found in Refs. 12 and 13.

A. Polymer graph and partition function

As shown in Fig. 3, we represent chain conformations as polymer graphs: chain monomers are represented by vertices, covalent bonds are represented by straight line links, and spatial contacts between monomers are represented by curved links.[12] Figure 3 shows a hairpin conformation and the corresponding polymer graph. A polymer graph is just a graphical form of the intrachain contact map. The polymer graph is a completely general representation, so our graph theory approach is not restricted to lattices or other simplified chain representations. The advantage for using the graphical representation is its mathematical convenience and simplicity, and its potential poser to analytically treat complicated chain conformations. A given polymer graph represents an ensemble of chain conformations that are consistent with the given contact constraints, but conformations having constants other than those specified by the polymer graph belong to other graphs. Any two pairs of curved links in a polymer graph must bear three types of relationships: nested, unrelated, or crossing linked; see Fig. 4. Graphs that involve no crossing linked links describe a large class of chain

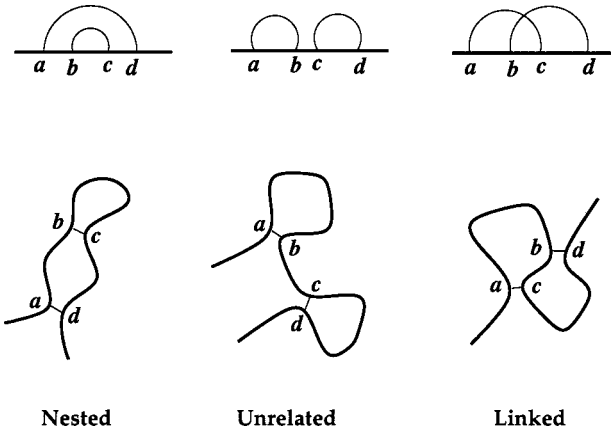

FIG. 4. Any two pairs of links must bear three types of relationships: nested, unrelated, or linked.

conformations—the double stranded chain conformations, examples of which include protein β-sheet structures and RNA secondary structures. In this work, we develop a theory for the calculation of the free energy landscapes for complexes of double stranded chain molecules.

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

$$Q_1 = \Sigma_{\text{graph}} \Omega e^{-E/kT}, \tag{4}$$

where E is the energy and Ω is the number of chain conformations that satisfy the intermonomer contact constraints depicted by the polymer graph. The key problems for the graphical approach are (1) to calculate Ω and (2) to sum over all the possible graphs (Σ_{graph}). To calculate Ω, we need to use a matrix method.

B. The matrix method for the number of chain conformations for a given polymer graph

Central to the computation of Ω is the excluded volume effect, i.e., two monomers cannot occupy the same position in space. Our approach to the excluded volume effect is ''dividing and conquering'' on the polymer graph.[12,13] By dividing a complex graph into (graphical) subunits, we can reduce a complex graph into computationally tractable smaller subsystems. A subunit is defined as a region in the graph bounded by the edges (either the straight or curved links) and containing no other edges in its interior (see Fig. 3, where a subunit is represented as a shaded region on the graph). Specifically, the calculation for Ω for a given graph involves two steps: to count all the viable chain conformations for each subunit, as if they are isolated and independent of each other, and to assemble the subunit conformations into the full conformations of the polymer graph. To assure that conformations of different subunits do not bump into each other (the excluded volume effect), a matrix method is used.

We classify the conformations of each subunit according to the shape of the ports through which it is connected to other subunits. On a 2D lattice, we classify the subunit conformations into four types.[12] To account for the interaction between the subunits, we define two matrices: (a) the structure matrix **S**, where matrix element S_{mn} in the number of

subunit chain conformations in which the ''input'' and ''output'' connections for the subunit of a graph are in the (mth, nth) type conformational state, $1 \leqslant m$, $n \leqslant 4$ on the 2D lattice, and (b) the viability matrix $\mathbf{Y}$, where elements $y_{mn}=1$ or 0 if connection between type n and m confirmations is viable or not viable.

A unique property of the polymer graphs for double-stranded chain conformations is that, through the graphical isomorphic transformations,[27] the graphs can be treated as or transformed into sequentially connected (nested) subunits, much like the way sections of train tracks or firehose are connected together.[12,13] Consequently, considering only excluded volume interactions between neighboring subunits, we can compute the number of chain conformations of the graph as a product of $\mathbf{S}$ and $\mathbf{Y}$ matrices for the subunits. For example, using the polymer graph in Fig. 3 as an illustration, the number of chain conformations for the graph closed by a curved link between vertex x and y is given by $\mathbf{U}\cdot\mathbf{\Omega}[x,y]\cdot\mathbf{U}^{\dagger}$, where $\mathbf{U}=[1,1,1,1]$ and $\mathbf{U}^{\dagger}$ is the transpose of $\mathbf{U}$, and matrix $\mathbf{\Omega}[x,y]$ is given by[12,13]

$$\mathbf{\Omega}[x,y]=\mathbf{S}^{(N)}\cdot\mathbf{Y}\cdot\mathbf{S}^{(N-1)}\cdot\mathbf{Y}\cdots\mathbf{S}^{(1)}, \tag{5}$$

where N is the number of (nested) subunits in the graph, $\mathbf{S}^{(j)}$ is the structure matrix of the jth subunit. To obtain the number of chain conformations for the whole graph [Ω in Eq. (4)], we need also to include the tails (x_0 to x and y to y_0 in Fig. 3), with the excluded volume effect properly considered.[13]

C. The partition function

To compute the partition function using Eq. (4), we need to sum over all the possible graphs [Σ_{graph} in Eq. (4)]. For the double stranded chain conformations, we can develop an efficient dynamic programming algorithm to calculate the sum.[13] The idea is to start with a short chain segment and elongate the chain by adding one monomer for each step, and to calculate recursively the partition function of the longer chain using the results for shorter chains.

We use the hairpin conformations in Fig. 3 for illustration. The partition function for the graph closed by the link between x and y can be calculated as $\mathbf{U}\cdot\mathbf{G}^{*}[x,y]\cdot\mathbf{U}^{\dagger}$, where matrix $\mathbf{G}^{*}[x,y]$ is defined as

$$\mathbf{G}^{*}[x,y]=\sum_{\text{all graphs}}\Omega[x,y]e^{-E/kT}$$

and according to Eq. (5), we can calculate $\mathbf{G}^{*}[x,y]$ through the following recursive relation:

$$\mathbf{G}^{*}[x,y]=\sum_{x',y'}\mathbf{S}^{(N)}\cdot\mathbf{Y}\cdot\mathbf{G}^{*}[x',y'], \tag{6}$$

where $\mathbf{S}^{(N)}$ is the structure matrix of the shaded subunit in Fig. 3.

The above matrix method is applicable to any double-stranded chain conformations. The method has been rigorously tested and validated for chain conformations on the 2D lattice,[13] and has been successfully applied to the prediction of free energy landscape for RNA secondary structures.[14]

IV. THE PARTITION FUNCTION OF COMPLEXES

In this paper, we develop a graph theoretic approach to the calculation of the partition function (Q_2) of the complexes (S_2). We represent conformations of the complex by a *generalized polymer graph*; see Fig. 5. We use two straight lines to represent the covalent bonds of the two chain molecules, align them in a straight line, and represent intrachain and interchain contacts as curved links connecting the corresponding monomers. Since we are limited to the double stranded conformations, we do not allow crossing links in the generalized polymer graph. The partition function of the complex can be defined as the weighted sum over all the possible generalized polymer graphs,

$$Q_2=\Sigma_{\text{generalized polymer graph}}\Omega e^{-E/kT}, \tag{7}$$

where E is the energy arising from the interchain and intrachain interactions, Ω is the number of complex conformations for the generalized polymer graph. We now develop a new graphical method to treat the conformations of complexes.

Because in the conformational space, spatial contacts (curved links in the graph) and covalent bonds (straight line links in the graph) constrain the chain configuration in very much the same way, we can deform the graph by changing the curved links into straight line links. For the generalized polymer graph, we apply this graphical transformation to the innermost curved link that join the two chains, for example, the link between monomer s_i and s_j' in Fig. 5. The transformation reduces the original generalized polymer graph into a composite graph consisting of two subgraphs—graphs A and B in Fig. 5. Because subgraphs A and B are polymer graphs for double-stranded conformations, they can be readily treated by the matrix method for a free molecules. We note that both A and B involve *intra*chain contacts, but subgraph B contains no *inter*chain contacts; all the interchain contacts are contained in subgraph A.

Since graphs A and B are divided by the innermost interchain link between s_i and s_j' in Fig. 5, we call the innermost interchain link a divider of the generalized polymer graph. The basic strategy of our approach is to classify the generalized polymer graphs according to the divider, compute the conditional partition function Q_2^{ij} for each given divided (s_i,s_j'), then the total partition function Q_2 of the complex, which involves the sum over all the possible generalized polymer graphs in Eq. (7), can be calculated as the sum of Q_2^{ij} over all the possible dividers,

$$Q_2=\sum_{i,j}Q_2^{ij}. \tag{8}$$

Subgraphs A and B can be treated as the polymer graphs of two chain molecules: chain A $(s_1s_2\cdots s_{i-1}s_is_j's_{j+1}'s_{j+2}'\cdots s_{L'}')$, and chain B $(s_Ls_{L-1}\cdots s_{i+1}s_is_j's_{j-1}'s_{j-2}'\cdots s_1')$; see Fig. 5. Chain A and B both contain a noncovalent bond—the link between s_i and s_j', and thus are not physical chains that consist of only covalently bonded monomers. We call chain A and B virtual chains.

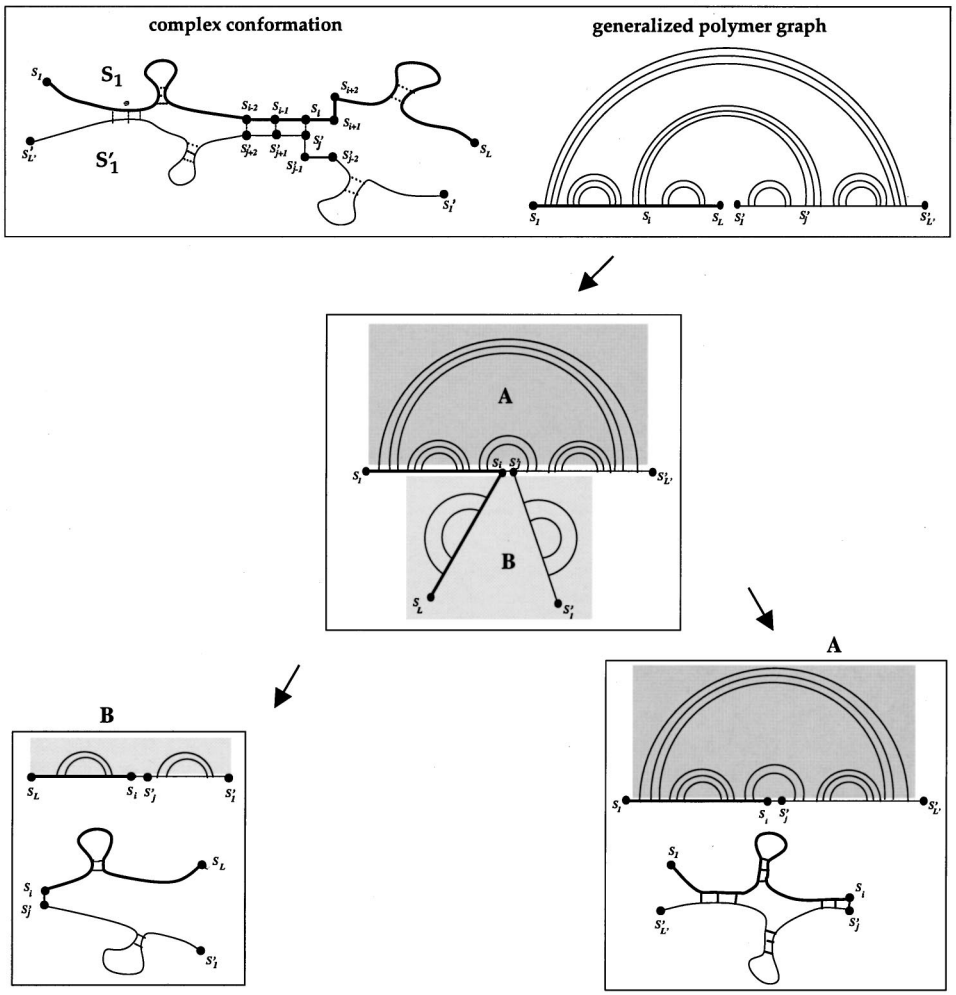

FIG. 5. The conformation and generalized polymer graph for a bound complex S_2 consisting of two chains: $S_1(s_1s_2\cdots s_{i-1}s_is_{i+1}\times s_{i+2}\cdots s_L)$ and $S_1'(s_1's_2'\cdots s_{j-1}'\times s_{j-1}'s_j's_{j+1}'s_{j+2}'\cdots s_{L'}')$. By treating the innermost interchain link between monomer s_i of chain S_1 and monomer s_j' of chain S_1' as an equivalent ''covalent bond'' (straight line) link, we can transform the original generalized polymer graph into two subgraphs A and B joined by the s_is_j' link.

To compute Q_2^{ij}, the partition function of the complex, we first compute the partition functions Q_A^{ij} and Q_B^{ij} of virtual chains A and B separately, treated them as isolated and independent chains. One might take the direct product $Q_A^{ij}\cdot Q_B^{ij}$ as an approximation for Q_2^{ij}. This is a good approximation only if A and B are disjoint graphs, corresponding to two independent virtual chains. However, since A and B are connected through the divider, they are dependent on each other through, for example, the excluded volume interactions. Therefore, $Q_A^{ij}\cdot Q_B^{ij}$ overestimates Q_2^{ij}. To account for the excluded volume interactions between A and B, we introduce a correlation factor Γ_{ij}, such that

$$Q_2^{ij}=Q_A^{ij}\cdot\Gamma_{ij}\cdot Q_B^{ij}. \tag{9}$$

In what follows, we develop a method to calculate partition functions Q_A^{ij}, Q_B^{ij} and correlation factor Γ_{ij}.

A. The partition function of the virtual chains

In this section, we calculate Q_A^{ij} and Q_B^{ij} for the virtual chains. The partition function Q_A^{ij} for chain A is calculated as the Boltzmann weighted sum of all the possible polymer graphs that involve noncrossing links. We directly apply the matrix method (see Sec. III and Refs. 12 and 13) developed for the double-stranded free molecules to the $(i-j+L'+1)$-mer chain A with the given sequence $s_1s_2\cdots s_{i-1}s_is_j's_{j+1}'s_{j+2}'\cdots s_{L'}'$. Note that part of the virtual chain belongs to the (physical) chain $S_1(s_1s_2\cdots s_{i-1}s_i)$, and the other part belongs to the (physical) chain $S_1'(s_j's_{j+1}'s_{j+2}'\cdots s_{L'}')$. Therefore, the partition function of virtual chain A contains the contributions from both interchain and intrachain contacts. Here interchain contacts refer to the contacts between monomers belonging to different (physical) chains (S_1 and S_1'), and intrachain contacts are the contacts between monomers belonging to the same (physical) chain.

Virtual chain B is a $(j-i+L+1)$-mer chain with sequence $s_Ls_{L-1}\cdots s_{i+1}s_is_j's_{j-1}'s_{j-2}'\cdots s_1'$. Chain B does not contain any inter-chain contacts, so the partition function Q_B^{ij} involves the sum over all the possible polymer graphs that contain only intrachain contacts. To calculate Q_B^{ij}, we need to modify the matrix method to exclude the interchain contacts. This can be readily done by setting $\mathbf{G}^*[x,y]=0$ [see Eq. (6) and Fig. 3] for any interchain contacts between monomer x and monomer y, i.e., for $x\in s_1,s_2,\ldots,s_L$ and $y\in s_1',s_2',\ldots s_{L'}'$, or $x\in s_1',s_2',\ldots,s_{L'}'$, and $y\in s_1,s_2,\ldots s_L$.

B. The correlation factor

The correlation factor, Γ_{ij} is introduced to account for the excluded volume interactions between subgraphs A and B. According to Eq. (9), Γ_{ij} is defined as

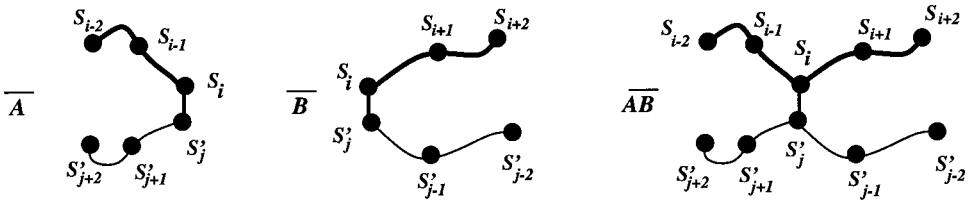

FIG. 6. Conformation of the reduced virtual chains $\overline{A}$, $\overline{B}$, and complex $\overline{AB}$ in the oligomeric system used to compute the correlation factor Γ_{ij}.

$$\Gamma_{ij}=Q_2^{ij}/\{Q_A^{ij}Q_B^{ij}\}. \tag{10}$$

$\Gamma_{ij}=1$ if (virtual) chain A and (virtual) chain B are completely independent, and $\Gamma_{ij}\ll 1$ if the steric conflictions between chain A and chain B are strong. Because B disallows any interchain contact, which effectively leads to more extended conformations, conformations of B would be more likely to bump into conformations of A. Therefore, we expect the excluded volume interactions to be strong and hence Γ_{ij} to be small in general.

Because the two (virtual) chains are joined at the divider, the excluded volume interactions would be dominant near the divider. Therefore, we focus on the monomers near the divider. To calculate Γ_{ij}, we reduce the complex AB to an oligomeric system that only includes monomers that are the nearest and next-nearest neighbors of monomers s_i and s'_j. Such a model system contains two reduced virtual chains $\overline{A}(s_{i-1}s_{i-1}s_is'_js'_{j+1}s'_{j+2})$ and $\overline{B}(s_{i+2}s_{i+1}s_is'_js'_{j-1}s'_{j-2})$; see Fig. 6. As a comparison, for free molecules, including the nearest neighbors would be sufficient to account for the excluded volume between subunits on the polymer graphs, but for the complexes, we need also to include the next nearest neighbors because the excluded volume confliction between A and B is stronger due to the more expanded conformations of B.

In the following, we apply Eq. (10) to the oligomeric system in Fig. 6 to derive Γ_{ij}. In this paper, we use the 2D lattice model for illustration, though the same methodology is applicable to more sophisticated chain representations. According to Eq. (10), Γ_{ij} can be determined by the following equation:

$$\Gamma_{ij}=Q_{\overline{AB}}^{ij}/\{Q_{\overline{A}}^{ij}Q_{\overline{B}}^{ij}\}, \tag{11}$$

where $Q_{\overline{A}}^{ij}$, $Q_{\overline{B}}^{ij}$, and $Q_{\overline{AB}}^{ij}$ are the partition functions for (virtual) chains $\overline{A}$, $\overline{B}$, and the complex $\overline{AB}$ in Fig. 6, respectively. The three partition functions can be calculated rigorously by exact enumeration on the 2D lattice.

$Q_{\overline{A}}^{ij}$ is calculated by enumerating all the possible double stranded conformations for the 6-mer chain $(s_{i-1}s_{i-1}s_is'_js'_{j+1}s'_{j+2})$ on the 2D lattice. There are totally 47 conformations for $\overline{A}$, including 29 conformations that involve no contact. We assume energy $E=0$ for conformations involving no any contact. The energy for a conformation with contacts depends on the sequence and energy function. There are totally $47-29=18$ conformations that involve contacts. These conformations are the first 18 conformations listed in Fig. 7. We denote the energy of the nth conformation in Fig. 7 by $E_n(n=1,2,\ldots,18)$, and define

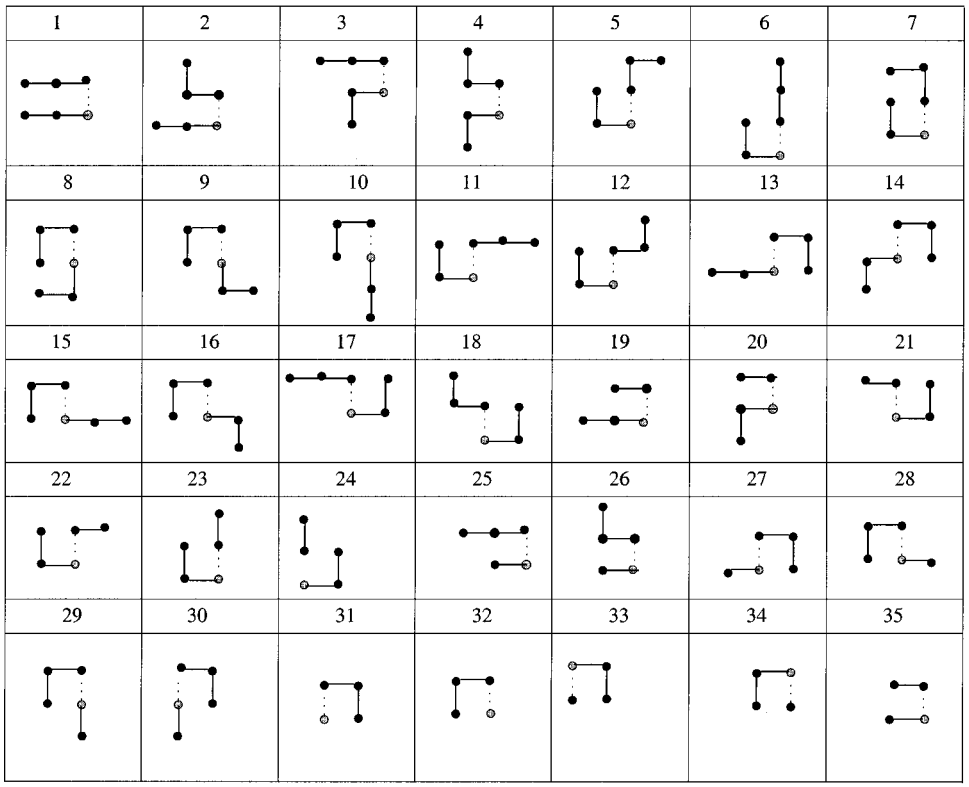

FIG. 7. The conformations of the reduced virtual chain $\overline{A}$ with different chain lengths (see Fig. 6). Only conformations involving contacts are included, because these are the only possible conformations that have nonzero energies. The dashed lines denote the links between s_i and s'_j, and the monomers in gray is s'_j.

TABLE I. The density of states of $\overline{AB}$.

Graph (n)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$g(E_n)$	24	12	4	4	0	0	2	2	2	0	2	2	0	2	0	2	0	2	0

$\omega_n = e^{-E_n/kT}$. Summing over all the conformations with and without contacts, we obtain the partition function $Q_{\overline{A}}^{ij}$,

$$Q_{\overline{A}}^{ij} = \sum_{n=1}^{18} \omega_n + 29.$$

The conformations that contribute to the partition function $Q_{\overline{B}}^{ij}$ do not contain any interchain contact. We find that the 29 conformations without contacts are the only viable conformation for chain $\overline{B}$. Therefore, the partition function of $\overline{B}$ is $Q_{\overline{B}}^{ij} = 29$.

To calculated $Q_{\overline{AB}}^{ij}$, the partition function for the complex $\overline{AB}$, we exhaustively enumerate all the viable conformations for the complex $\overline{AB}$. Since none of the conformations of $\overline{B}$ contains contact, the energy of $\overline{B}$ is zero, therefore, the energy of $\overline{AB}$ comes from the contribution of $\overline{A}$. By collecting up conformations according to their energies, we obtain the density of states $g(E)$ for the complex, as shown in Table I. With the density of states, we can write $Q_{\overline{AB}}^{ij}$ as

$$Q_{\overline{AB}}^{ij} = \sum_{n=0}^{18} g(E_n) e^{-E_n/kT},$$

where $n=0$ denotes the conformations without any contact. From the above results for $Q_{\overline{A}}^{ij}$, $Q_{\overline{B}}^{ij}$, and $Q_{\overline{AB}}^{ij}$, we can obtain Γ_{ij} from Eq. (11).

The above derivation for Γ_{ij} assumes the $\overline{A}$ and $\overline{B}$ both are 6-mer chain segments, which occurs for $3 \leqslant i \leqslant L-2$ and $3 \leqslant j \leqslant L'-2$. However, when i or j is close to either end of the chains, $\overline{A}$ or $\overline{B}$ may becomes shorter than 6-mer, in which case we must compute Γ_{ij} separately. For example, if $i = 1, 3 \leqslant j \leqslant L'-2$, then $\overline{A}$ becomes a 5-mer chain. In Fig. 7, for different i and j's, we exhaustively list all the possible $\overline{A}$ conformations that contain contacts. These are the only possible conformations that have nonzero energy contributions to $\overline{A}$ and $\overline{AB}$. For example, the conformations for a 5-mer chain are listed in the 19th to 24th item. We denote the energy for the nth conformation in Fig. 7 to be E_n, and defined $\omega_n = e^{-E_n/kT}$. In Table II, for different i and j's we list the results in terms of ω_n for $Q_{\overline{A}}^{ij}$, $Q_{\overline{B}}^{ij}$, and $Q_{\overline{AB}}^{ij}$, from which Γ_{ij} can be determined by Eq. (11).

C. Test of the model

The most direct test is for system with the simplest energy function. To test our theory, we choose a model systems consisting of two homopolymer chains, and assume that monomers in contact stick to each other with an energy $-\epsilon$. We use the above theory to compute the partition functions for complexes of short homopolymer chains at different temperatures, and compared the results to those obtained from exact computer enumeration.

We first validate the calculation for the partition function Q_2^{ij} for fixed positions of the divider (i,j) [see Eq. (8)]. Figure 8(A) shows the comparison of the theory and the exact computer enumeration for a complex of two 12-mer chains with $(i,j)=(11,2)$ and (3,10). Our results show perfect agreement for a wide temperature range. We then test for the total partition function of the complex Q_2 by summing over all the possible i and j's [see Eq. (8)]. We have tested for different chain length $(L,L')=(12,12)$ and (13,6). As shown in Fig. 8(B), the comparisons with exact computer enumeration again show that our theoretical predictions are essentially exact.

V. PREDICTING FREE ENERGY LANDSCAPES OF COMPLEXES

In this section, we illustrate the applications of the theory by studying the free energy landscapes and the conformational change as a function of temperature for several 2D lattice model systems. The purpose of the calculation is to demonstrate the generality of the model to treat homopolymers and heteropolymers with specific sequences and with different energy functions, and to investigate the free energy landscapes of these systems.

To investigate the structural transitions as temperature is changed, we calculated the heat capacities $C_1(T)$, $C_1'(T)$, and $C_2(T)$ of free molecules S_1, S_1', and the complex S_2, respectively. Heat capacities are determined by the second order derivatives of the partition functions and the free energies.[28] For example, $C_2(T) = (\partial/\partial T)[kT^2(\partial/\partial T)\ln Q_2]$, where $Q_2 = \Sigma_{\mathbf{x}} Q_2(\mathbf{x})$ is the total partition function for all the possible conformations $\mathbf{x}$ of the complexes. The peaks of the melting curves suggest structural transitions. Though $C_2(T)$, $C_1(T)$, and $C_1'(T)$ are not the heat capacities of the total system $(S_1+S_1'+S_2)$ that involve free molecules S_1, S_1', and bound complex S_2, as we still see, the heat capacity melting curves facilitate the analysis of the conformational transitions.

To compute the free energy landscapes $\Delta F(\mathbf{x})$, defined in Eqs. (1) and (3), we must first choose the parameter $\mathbf{x}$ used to describe the chain conformations. In order to depict free energy landscapes as useful 3D graphical representations, we need to plot $\Delta F(\mathbf{x})$ as a function of *two* variables. To directly relate the free energy landscape to the association and dissociation of the complex, we choose $\mathbf{x}=(N,NN)$, where N is the total number of intrachain contacts and NN is the number of interchain contacts. $NN \geqslant 1$ corresponds to the state of a complex, and $NN=0$ corresponds to the state of separated chains (free molecules). Each point (N,NN) represents a large ensemble of conformations that are collected together under the same number of intrachain and interchain contacts, and the thermodynamic average of these conformations gives $\Delta F(N,NN)$.

TABLE II. Partition functions $Q_{\overline{AB}}^{ij}$, $Q_{\overline{A}}^{ij}$, and $Q_{\overline{B}}^{ij}$, for the virtual chains $\overline{A}$, $\overline{B}$, and complex $\overline{AB}$, respectively. With the partition functions, Γ_{ij} can be determined from Eq. (11).

i,j	$Q_{\overline{A}}^{ij}$	$Q_{\overline{B}}^{ij}$	$Q_{\overline{AB}}^{ij}$
$3\leqslant i\leqslant L-2$, $3\leqslant j\leqslant L'-2$	$\Sigma_{i=1}^{18}\omega_i+29$	29	$12\omega_1+4\omega_2+4\omega_3+2(\omega_6+\omega_7+\omega_8+\omega_{10}+\omega_{11}+\omega_{13}+\omega_{15}+\omega_{17})+24$
$i=2$, $3\leqslant j\leqslant L'-2$	$\Sigma_{i=19}^{24}\omega_i+12$	29	$9\omega_{19}+2\omega_{20}+\omega_{21}+\omega_{22}+\omega_{23}+\omega_{24}+17$
$j=L'-1$, $3\leqslant i\leqslant L-2$	$\Sigma_{i=25}^{30}\omega_i+12$	29	$9\omega_{25}+2\omega_{26}+\omega_{27}+\omega_{28}+\omega_{29}+\omega_{30}+17$
$i=L-1$, $3\leqslant j\leqslant L'-2$	$\Sigma_{i=1}^{18}\omega_i+29$	12	$6\omega_1+2(\omega_2+\omega_3+\omega_8+\omega_{10}+\omega_{13}+\omega_{15})+\omega_6+\omega_7+\omega_{11}+\omega_{17}+18$
$j=2$, $3\leqslant i\leqslant L-2$	$\Sigma_{i=1}^{18}\omega_i+29$	12	$6\omega_1+2(\omega_2+\omega_3+\omega_6+\omega_7+\omega_{11}+\omega_{17})+\omega_8+\omega_{10}+\omega_{13}+\omega_{15}+18$
$i=1$, $3\leqslant j\leqslant L'-2$	$\omega_{33}+\omega_{34}+7$	29	$12\omega_{33}+12\omega_{34}+48$
$j=L'$, $3\leqslant i\leqslant L-2$	$\omega_{31}+\omega_{32}+7$	29	$12\omega_{31}+12\omega_{32}+48$
$i=L$, $3\leqslant j\leqslant L'-2$	$\Sigma_{i=1}^{18}\omega_i+29$	7	$4(\omega_1+\omega_2+\omega_5+\omega_6+\omega_7)+2(\omega_3+\omega_4+\omega_{11}+\omega_{12}+\omega_{17}+\omega_{18})+\omega_8+\omega_{10}+\omega_{13}+\omega_{15}+32$
$j=1$, $3\leqslant i\leqslant L-2$	$\Sigma_{i=1}^{18}\omega_i+29$	7	$4(\omega_1+\omega_3+\omega_8+\omega_9+\omega_{10})+2(\omega_2+\omega_4+\omega_{11}+\omega_{12}+\omega_{17}+\omega_{18})+\omega_8+\omega_{10}+\omega_{13}+\omega_{15}+32$
$i=L-1$, $j=1$	$\Sigma_{i=1}^{18}\omega_i+29$	3	$2(\omega_1+\omega_3+\omega_8+\omega_9+\omega_{10})+\omega_2+\omega_4+\omega_6+\omega_7+\omega_{11}+\omega_{12}+\omega_{13}+\omega_{14}+\omega_{16}+\omega_{17}+30$
$i=L$, $j=2$	$\Sigma_{i=1}^{18}\omega_i+29$	3	$2(\omega_1+\omega_3+\omega_5+\omega_6+\omega_7)+\omega_2+\omega_4+\omega_8+\omega_{10}+\omega_{11}+\omega_{12}+\omega_{13}+\omega_{14}+\omega_{15}+\omega_{17}+30$
$i=2$, $j=L'$ or $i=1$, $j=L'-1$	3	29	29
$i=2$, $j=L'-1$	$\omega_{35}+5$	29	$16\omega_{35}+14$
$i=L-1$, $j=2$	$\Sigma_{i=1}^{18}\omega_i+29$	5	$3\omega_1+\omega_2+\omega_3+\omega_6+\omega_7+\omega_8+\omega_{10}+\omega_{11}+\omega_{13}+\omega_{15}+\omega_{17}+12$
$i=2$, $j=2$	$\Sigma_{i=19}^{24}\omega_i+12$	12	$6\omega_{19}+2\omega_{20}+2\omega_{21}+2\omega_{22}+\omega_{23}+\omega_{24}+12$
$i=L-1$, $j=L'-1$	$\Sigma_{i=25}^{30}\omega_i+12$	12	$6\omega_{25}+2\omega_{26}+2\omega_{27}+2\omega_{28}+\omega_{29}+\omega_{30}+12$
$i=1$, $j=1$	$\omega_{33}+\omega_{34}+7$	7	$4\omega_{33}+4\omega_{34}+34$
$i=L$, $j=L'$	$\omega_{31}+\omega_{32}+7$	7	$4\omega_{31}+4\omega_{32}+34$
$i=L-1$, $j=L'$	$\omega_{31}+\omega_{32}+7$	7	$5\omega_{31}+5\omega_{32}+27$
$i=1$, $j=2$	$\omega_{33}+\omega_{34}+7$	7	$5\omega_{33}+5\omega_{34}+27$
$i=L$, $j=L'-1$	$\Sigma_{i=25}^{30}\omega_i+12$	7	$4\omega_{25}+4\omega_{26}+\omega_{27}+\omega_{28}+\omega_{29}+\omega_{30}+28$
$i=2$, $j=1$	$\Sigma_{i=19}^{24}\omega_i+12$	7	$4\omega_{19}+4\omega_{20}+\omega_{21}+\omega_{22}+\omega_{23}+\omega_{24}+28$

Change of temperature causes the change of the shape of the free energy landscape, and thus the changes of the stable states of the system. To obtain the structural information from the free energy landscapes, we calculate the contact probability $p(x,y)$ for each pair of monomers (x,y). Since an intrachain contact pair between monomer x and y could occur either in the complex or in a free molecules, we compute $p(x,y)$ for complexes and for free molecules separately. $p(x,y)$ is calculated from the partition functions,[29]

$$p(x,y)=Q_2(x,y)/Q_{\text{tot}} \quad \text{for the complex,}$$

$$p(x,y)=Q_1(x,y)\cdot Q_2/Q_{\text{tot}} \quad \text{for the free molecule } S_1,$$

$$p(x,y)=Q_1'(x,y)\cdot Q_2/Q_{\text{tot}} \quad \text{for the free molecules } S_1',$$

where $Q_2(x,y)$, $Q_1(x,y)$, and $Q_1'(x,y)$ are the conditional partition functions for S_2, S_1, and S_1', respectively, under the condition that monomers x and y must form a contact, Q_2, Q_1, and Q_1' are the partition functions of the complex S_2 and of free molecule S_1 and S_1', respectively, Q_{tot} is the total partition function, $Q_{\text{tot}}=Q_2+\alpha Q_1Q_1'$. From the contact probabilities, we can deduce the structures or the structural distributions.

We represent $p(x,y)$ by a density plot. For an L-mer chain S_1 and L'-mer chain S_1', the density plot is a $[L+L']\times[L+L']$ matrix, where the first L columns and rows represent the L monomers of molecule S_1, and the next L' columns and rows represent the monomers of molecule S_1'. We use the upper half $(y>x)$ of the plot for the contacts in the complex S_2, and the lower half $(y<x)$ for the contacts

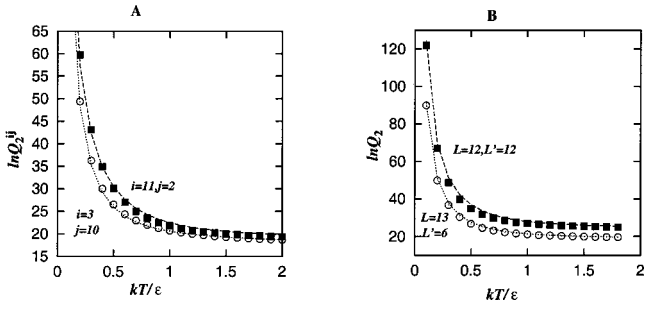

FIG. 8. (A) Tests for the partition function Q_2^{ij} for the complex of two 12-mer homopolymer chains, for $(i,j)=(3,10)$ and $(11,2)$. (B) Test for the total partition function Q_2 of the complex for different chain length $(L,L')=(12,12)$ and $(13,6)$. We assume monomers in contact stick to each other with an energy $-\epsilon$. The symbols are for the exact results from exhaustive computer enumeration on the 2D lattice, and the lines are for the results from our theory.

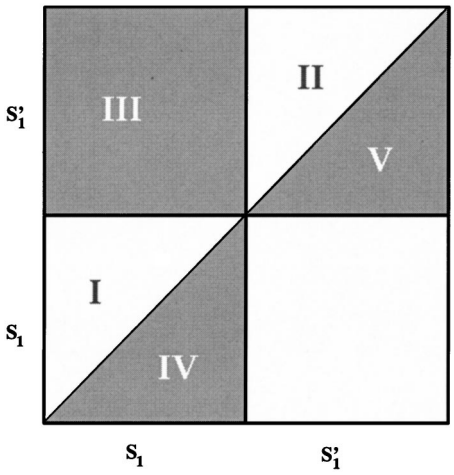

FIG. 9. The plot of the contact probability $p(x,y)$ is divided into five regions: (I) self-contacts of chain S_1 in the complex, (II) self-contacts of chain S_1' in the complex, (III) interchain contacts between chain S_1' and S_1' in the complex, (IV) self-contacts in the free chain S_1, and (V) self-contacts in the free chain S_1'. The first L rows and columns represent the sequence of chain S_1, and the next L' rows and columns represent the sequence of chain S_1'. L and L' are the length of the S_1 and S_1', respectively.

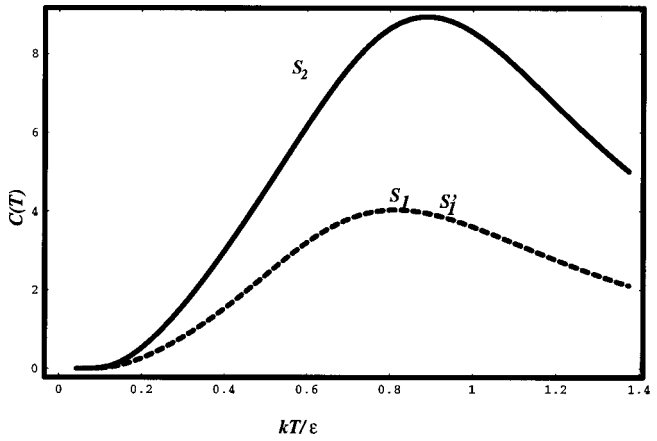

FIG. 10. The predicted melting curve for the heat capacities of the complex S_2 of two 16-mer homopolymer and the free homopolymer S_1. We assume monomers in contact stick to each other with an energy $-\epsilon$.

for the free molecules. As shown in Fig. 9, in this way, we can divide the density plot into five regions, corresponding to the probability for the formation of five different types of contacts: the self-contact of chain S_1 and of chain S_1' in the bound complex and in the free molecules, and the interchain contacts in the complex.

Our theory is applicable to the complexes of chains of any length and any sequence, for any given energy functions (Hamiltonians). In the following, we study the formation of homopolymer–homopolymer, heteropolymer–homopolymer, and heteropolymer–heteropolymer complexes, with increasing sequence heterogeneity. We choose 16-mer chains for both S_1 and S_1'. For convenience, we consider the melting of the complexes, rather than the annealing of two chains into a complex. For the equilibrium thermodynamics, the melting and annealing process are reversible.

A. Homopolymer–homopolymer complex with contact-based energy functions

We consider the association of two 16-mer homopolymers that have identical sequences: *AAAAAAAAAAAAAAAA*. We assume that each contact has an attractive energy $-\epsilon$.

We first calculate the heat capacity melting curves of the complex and for the free molecule. From the melting curves (see Fig. 10), we find that the complex and the free molecule show single peaked melting curves at $kT/\epsilon \simeq 0.8-0.9$, suggesting possible structural transitions near this temperature region.

The free energy landscapes and structural changes will certainly depend on the parameter α defined in Eq. (2). Dissociation of the complex becomes more favorable for larger α than for smaller α, and for α smaller than certain critical value, the dissociation of the complex may become impossible for a thermodynamic equilibrium system. We will discuss two representative values of α ($\alpha=18$ and $\alpha=5000$) separately, because they have distinctive energy landscape for the melting process. We show the density plot for the contact probability $p(x,y)$ and the 3D surface plot for the free energy landscape $\Delta F(N,NN)$ in Fig. 11 and Fig. 12 for $\alpha=18$ and $\alpha=5000$, respectively.

For $\alpha=18$ (Fig. 11), the $p(x,y)$ plot shows that a low temperatures (e.g., at $kT/\epsilon=0.1$), the homopolymers form complexes, and most of the complexes involve the end–end interchain contact between $x=1$ of S_1 and $y=16$ of S_1' or between $x=16$ of S_1 and $y=1$ of S_1'. Complexes are formed at low temperature because they can achieve the maximum number of (energetically favorable) contacts. There are a large ensemble of conformations of the complex that can have the maximum number of contacts. The free energy landscape shows that at $kT/\epsilon=0.1$, the free energy minima are continuously distributed along the diagonal $N+NN=16$, indicating an ''amorphous'' molten native structure of the complex. At higher temperatures, the system remains to be a complex state, but the interchain and intrachain contacts are disrupted, and the most favorable contacts are the local self-contacts between monomers x and $y=x+3$ ($1\leqslant x \leqslant 13$), because these local contacts cause minimal entropic loss. When $kT/\epsilon \geqslant 1$, the global minimum on the landscape is on $NN=1$, corresponding to the complex with only one interchain contact, and the minimum continuously moves along $NN=1$ to small N values, corresponding to the disruption of the intrachain contacts.

An intriguing phenomena for the homopolymer system with $\alpha=18$ is the consistent dominance of the complex structures throughout the melting process. In other words, the two chains virtually never separate, even at high temperatures. The physical reason is the following. According to its definition, a complex involves at least one interchain contact. For a complex, different ways of forming the interchain contacts lead to a large combinatory entropy. When chains are separated, the system loses this combinatory entropy (of the complex), but gains the translational and rotational entropy (of free molecules). In order for the complete chain separation to occur, α must be at least large enough to compensate the loss of the combinatory entropy. There exists a critical

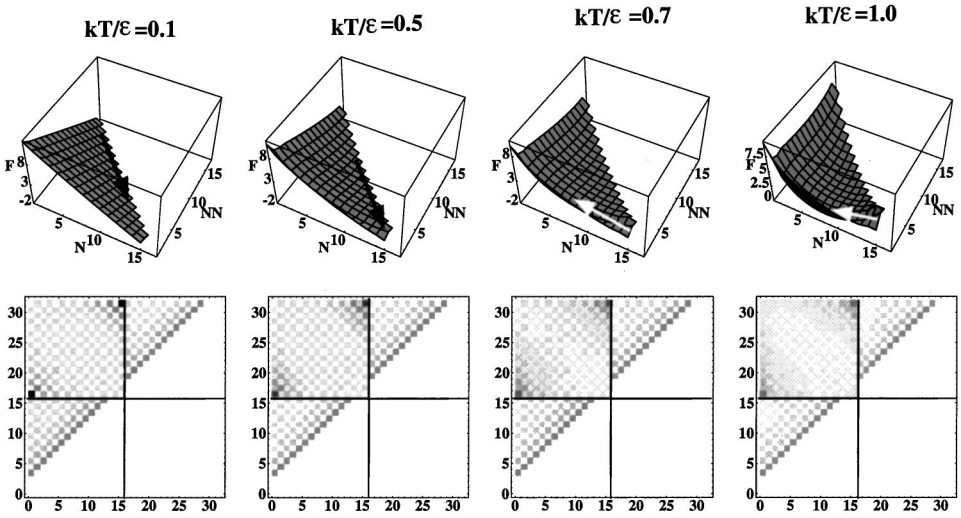

FIG. 11. The free energy landscape and contact probability at different temperatures for $\alpha=18$ of the two homopolymer complex.

value α^* in order for the complex to eventually dissociate (at high temperature),

$$\alpha^* = Q_2/[Q_1 Q_1']|_{T\to\infty}. \tag{12}$$

We found $\alpha^* \cong 500$ for the homopolymer–homopolymer system.

For $\alpha=5000(>\alpha^*)$ (Fig. 12), the chains form complexes at low temperatures ($kT/\epsilon=0.1$). As temperature is increased to $kT/\epsilon=0.5$, chain separation occurs. Interestingly, at the transition temperature, the contact probability $p(x,y)$ for the self-contacts of chain $S_1(S_1')$ in the complex and in the free molecules show nonsymmetric distributions, indicating that the dissociation of the complex involves redistribution of the chain conformations. The free molecules not only involve local contacts between monomers x and $y=x+3(1\leqslant x\leqslant 13)$ that exist in complexes but also have nonlocal contacts between monomer $x=1$ and $y=16$ because the nonlocal contact are well populated in the first few excited states.[13] As temperature is further increased (e.g., $kT/\epsilon=1$), two chains are completely separated, and the chains exist as free molecules.

The free energy landscape at low temperature ($kT/\epsilon=0.1$) has the same shape as that for $\alpha=18$. However, unlike the $\alpha=18$ case, for kT/ϵ larger than 0.5, besides the minima that correspond to the low energy molten state of the complex, the landscape shows a new deep minimum at $NN=0$, corresponding to a stable state of separated chains. These two sets of minima coexist, and the barrier between them shows that the dissociation of the complex is a thermodynamically cooperative transition. At higher temperatures, the global minimum of the landscape moves along the $NN=0$ line, corresponding to the unfolding of the free molecules.

B. Homopolymer–heteropolymer complex with contact-based energy functions

We next study the binding of a 16-mer heteropolymer and a 16-mer homopolymer chain. We assume a binary sequence: *BABABABAABABABAB* for S_1, and a homopolymer sequence: *AAAAAAAAAAAAAAAA* for S_1'. We define an attractive energy $-\epsilon$ for each $A-A$ or $B-B$ contact, and energy 0 for $A-B$ contacts.

The heat capacity melting curves of the complex S_2 and the free molecules S_1 and S_1' are shown in Fig. 13. We find that the complex undergoes two transitions at $kT/\epsilon\cong 0.3$ and

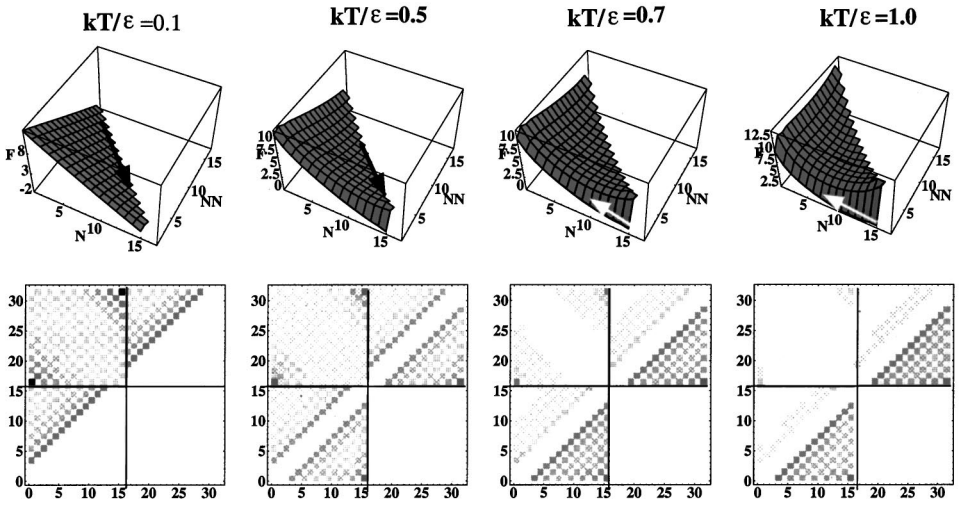

FIG. 12. The free energy landscape and contact probability at different temperatures for $\alpha=5000$ of the two homopolymer complex.

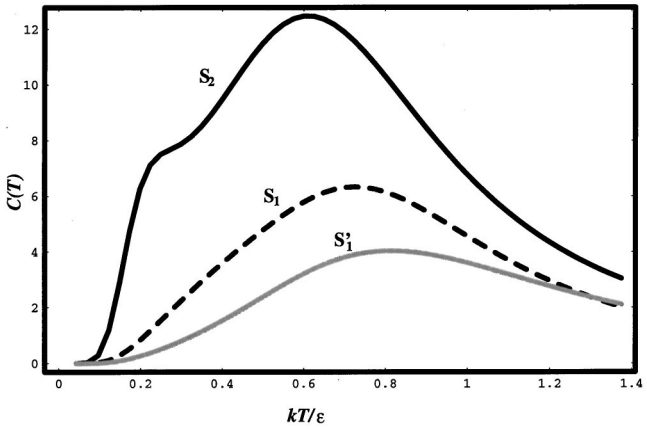

FIG. 13. The predicted melting curve for the heat capacities of a 16-mer heteropolymer (S_1), a 16-mer homopolymer (S_1'), and the bound complex S_2 of the homopolymer and the heteropolymer. Chain S_1 and S_1' have sequences *BABABABAABABABAB* and *AAAAAAAAAAAAAAAA*, respectively. We assume an attractive energy $-\epsilon$ for each A–A or B–B contact, and energy 0 for each AB contact.

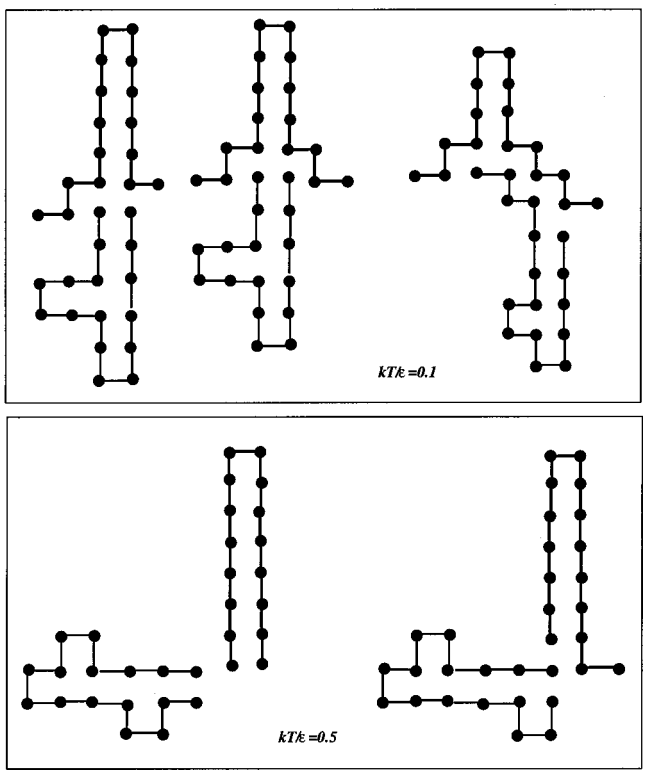

FIG. 15. Three typical structures of a 16-mer heteropolymer and 16-mer homopolymer bound complex at $kT/\epsilon=0.1$ (A) and at $kT/\epsilon=0.5$ (B).

0.6, respectively, and the free molecules S_1 and S_1' undergo single transitions at $kT/\epsilon \cong 0.7$ and 0.8, respectively.

Equation (12) gives the critical α value for the chain separation to occur: $\alpha^* \cong 85$ for the homopolymer–heteropolymer system. We discuss two cases $\alpha=4(<a^*)$ and $\alpha=500(>\alpha^*)$ separately.

Throughout the melting process, we find that, the system involves multiple structural transitions, and due to the presence of the homopolymer, each phase of the system is in an ''amorphous'' state with no uniquely defined structures.

1. $\alpha=4$

See Fig. 14 for the free energy landscape and the contact probability plot.

a. Noncooperative chain separation: At low temperature $kT/\epsilon=0.1$, the free energy landscape shows a series of minima along the diagonal $N+NN=15$ with extensive interchain contacts ($NN \geqslant 4$), suggesting a molten native state of the bound complex. Figure 15(A) shows three typical native structures obtained from the contact probability plot in Fig. 14. These structures contain the maximum number ($N+NN=15$) of the A–A contacts with the lowest energy $E=-15\epsilon$. As temperature is increased, the minima on the free energy landscape gradually deviate from the diagonal $N+NN=15$, and in the meantime move toward position with small NN values, suggesting the separation of the chains. Such a structural transition corresponds to the first peak at $kT/\epsilon \cong 0.3$ for the heat capacity melting curve of the complex in Fig. 13. Because the positions of the minima continuously change as T is changed, the conformational ''transition'' is noncooperative. Figure 15(B) shows two structures of the complex at $kT/\epsilon=0.5$. For $\alpha=4$, the complete chain separation ($NN=0$) does not occur, even at high tempera-

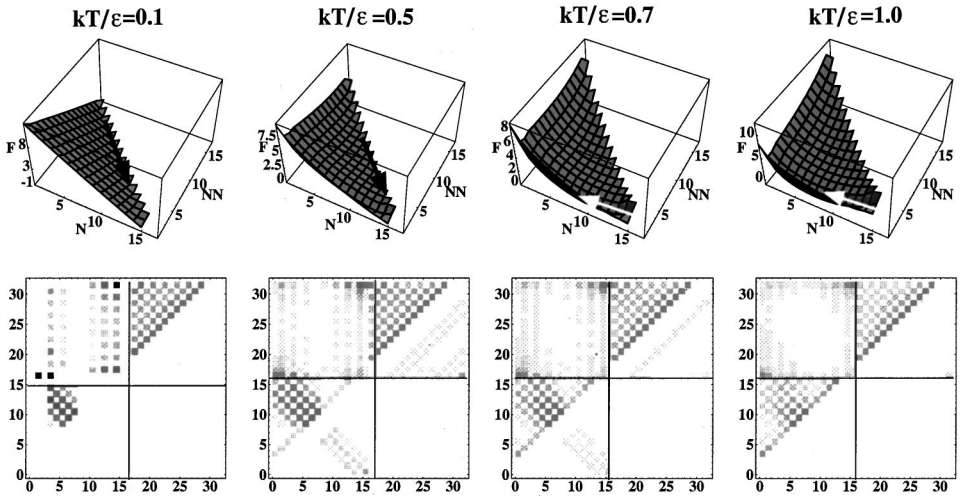

FIG. 14. The free energy landscape and contact probability at different temperatures for $\alpha=4$ of the heteropolymer–homopolymer system.

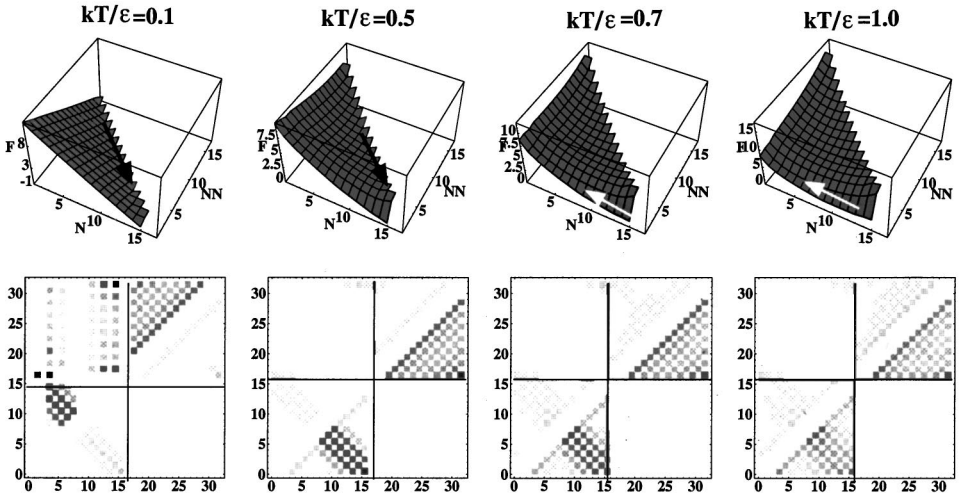

FIG. 16. The free energy landscape and contact probability at different temperatures for $\alpha=500$ of the heteropolymer–homopolymer system.

ture. As temperature T is further increased, the minima of the free energy landscape move to the smallest possible value of NN for a bound complex: $NN=1$. Complexes with $NN=1$ have the maximum entropy for given (N) intrachain contacts.

b. Noncooperative melting of the intrachain contacts: As temperature is further increased, the minima on the free energy landscape move along $NN=1$ toward smaller N values, corresponding to the melting of the intrachain contacts. As shown in Fig. 14, the minima move continuously, so the structural change is noncooperative.

2. $\alpha=500$

See Fig. 16 for the free energy landscape and the contact probability plot.

a. Cooperative chain separation: On the free energy landscape, the native state corresponds to a series of minima along the diagonal $N+NN=15$. Figure 15(A) shows three typical structures of the native state. As T is increased, the minima move continuously toward smaller $NN(NN\geqslant 1)$ values, corresponding to the gradual separation of the complex. As T is further increased, a new minimum at $NN=0$, $N=15$ appears, corresponding to a new stable state—the completely separated free molecules. The minima for the stable states of the complex and the new minimum for the free molecules are divided by a free energy barrier, indicating a cooperative transition for the complete dissociation of the complex. For larger α, the barrier is higher, and the chain separation transition is more cooperative.

b. Noncooperative melting of the intrachain contacts: After the complex is completely dissociated ($NN=0$), the contacts in the free chains will be disrupted as temperature is further increased. On the free energy landscape, this is manifested as the continuous move of the minimum along $NN=0$ toward smaller N values (see Fig. 16). Since minimum moves continuously, the structural change is noncooperative.

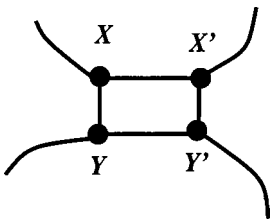

FIG. 17. The stack conformation formed by two consecutive contacts (X,Y) and (X',Y'). We assume four types of monomers: *A, U, C, G,* to mimic the RNA stacking interactions.

C. Heteropolymer–heteropolymer complex with stacking-based energy functions

The thermal melting transition of a heteropolymer–heteropolymer complex certainly depends on the sequences of the heteropolymers. We assume four-letter codes (*A, U, C,* and *G*) for the sequences of the heteropolymers to mimic the four types of nucleotides in RNA molecules. In our calculation, we choose sequence *AAAAAUUUUUCCGGGG* for chain S_1 and *AAAAAUUUUUGGCCCC* for chain S_1'. Both chains have 16 monomers. We use a RNA-like base stacking energy functions for the system (see Fig. 17). We assume that only two stacks formed by Watson–Crick pairs ($A-U$ or $C-G$) have nonzero energy, and each $A-U$ pair contributes an energy $-\epsilon/2$, and each $G-C$ pair contributes and energy $-\epsilon$.

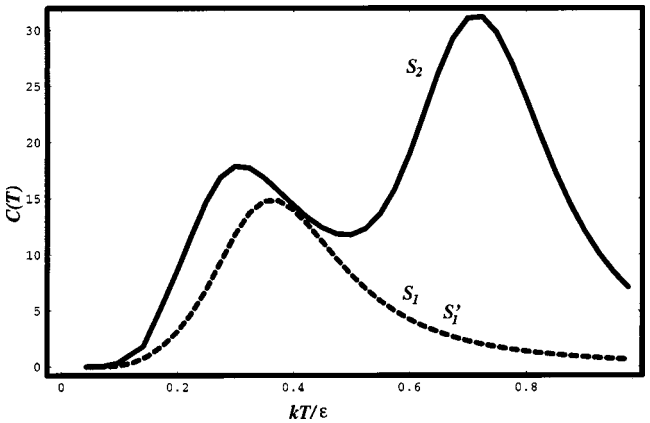

FIG. 18. The predicted melting curve for the heat capacities of two 16-mer heteropolymers: *AAAAAUUUUUCCGGGG* (S_1) and *AAAAAUUUUUGGCCCC* (S_1'), and the bound complex S_2 of these two heteropolymers. The stacking energy function is used. Interestingly, because of the complementarity between the sequence of S_1 and of S_1', they show the same melting curve.

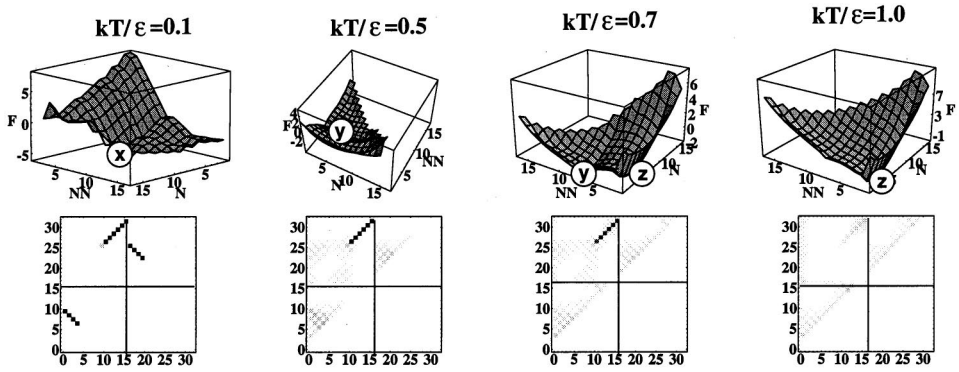

FIG. 19. The free energy landscape and contact probability at different temperatures for $\alpha=4$ of the heteropolymer–heteropolymer system.

We first calculate the heat capacity melting curves for the complex S_2 and the two free molecules S_1 and S_1'. As shown in Fig. 18, the melting curves of the complex has two peaks at $kT/\epsilon \cong 0.3$ and 0.7, respectively, suggesting two structural transitions during the melting process, and the free molecules involve single transitions.

The critical value of α for the chain separation to occur is estimated from Eq. (12): $\alpha^* \simeq 410$. Therefore, we discuss two different α's: $\alpha=4(<\alpha^*)$ and $\alpha=2000(>\alpha^*)$ separately. Figure 19 and Fig. 20 show the free energy landscapes and the contact probability $p(x,y)$ for $\alpha=4$ and $\alpha=2000$, respectively.

At low temperature $kT/\epsilon=0.1$, the free energy landscape has a global minimum X at position $(NN,N)=(6,8)$, which, according to the contact probability plot, corresponds to three native structures, as shown in Fig. 21. By comparing the native structures of the complex and the native structures of the free molecules in Fig. 21, we find that the complex S_2 is formed from S_1 and S_1' through induced fit because both S_1 and S_1' undergo significant structural rearrangement upon binding.

As temperature is increased, a new minimum (valley) at Y is formed on the landscape along $NN=6$ with $0 \leqslant N \leqslant 8$ (see Fig. 19), representing an ensemble of conformations of the partially dissociated complex that contain six interchain contacts. The intermediate state Y has a well-defined structure, as shown in Fig. 21 for $kT/\epsilon=0.5$. Y is formed by the melting of the intrachain contacts while keeping the interchain contact intact. This shows that the interchain contacts are more stable than the intrachain contacts. Physically this is because the $C-G$ bonds in the interchain contacts are energetically more stable than the $A-U$ bonds in the intrachain contacts in the native state X of the complex. The transition from the native state X to the intermediate state Y occurs at $kT/\epsilon \simeq 0.3$, corresponding to the first peak, of the melting curves of the complex in Fig. 18. Because X and Y are separated by a free energy barrier on the landscape, the transition from X to Y is thermodynamically cooperative.

As the temperature is further increased, at $kT/\epsilon \simeq 0.7$, another new valley (minimum) Z appears on the landscape at $NN=1$ for small α (e.g., $\alpha=4$), or at $NN=0$ for large α (e.g., $\alpha=500$). Z corresponds to the final denatured state of the system: $NN=1$ state corresponds to the denatured complex structure with only one interchain contact, and the $NN=0$ state corresponds to the completely separated chains. Since the Y and Z are separated by a barrier, the transition from the intermediate state Y to the final denatured state Z is cooperative.

V. SUMMARY

We have developed an analytical method to rigorously compute the free energy landscapes for double-stranded chain molecules. Our new model can include the complete ensemble of double-stranded conformations for the free molecules and the bound complexes. Therefore, this model can treat the conformational rearrangement upon binding of chain molecules (''induced fit''), and can treat free energy changes and binding affinities based on the full ensemble of chain conformations, including native and non-native states.

Our model studies for the homopolymer–homopolymer, homopolymer–heteropolymer, and heteropolymer–heteropolymer complex systems show quite distinctive free energy landscapes and structural changes for these systems,

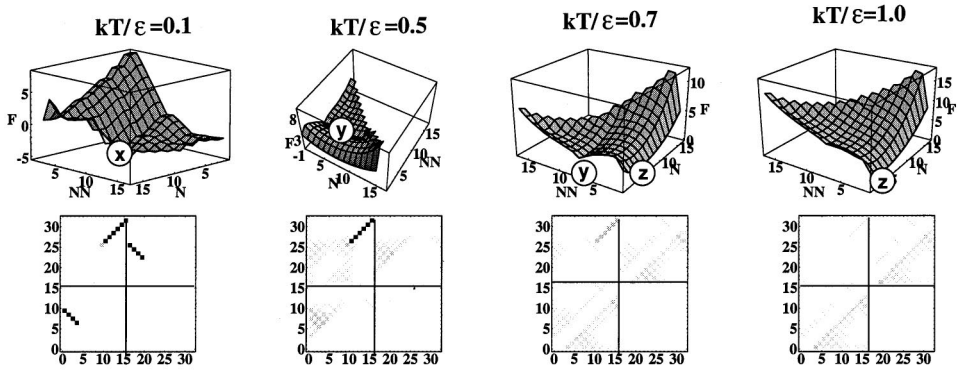

FIG. 20. The free energy landscape and contact probability at different temperatures for $\alpha=2000$ of the heteropolymer–heteropolymer system.

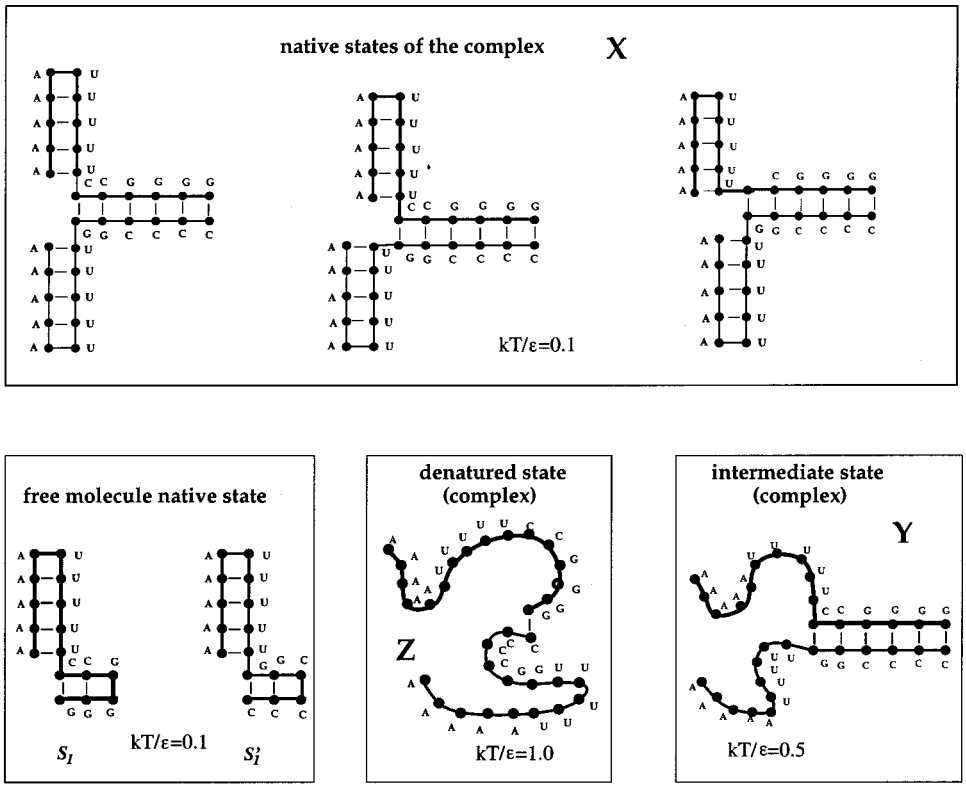

FIG. 21. The native states of the free molecules and stable structures of the heteropolymer–heteropolymer complex at different temperatures.

depending on the competition between the conformational free energy of the complex and the rotational and translational free energy of the free molecules. We find that the free energy landscapes for systems involving homopolymers are smoother than the heteropolymer systems, and the structural changes are less cooperative. For the heteropolymer system, we find that the landscapes are bumpy, involving intermediate states of bound complexes prior to complete dissociation, and the system involves cooperative structural transitions.

Though our theory was tested and validated on the 2D lattice, the general methodology is applicable to off-lattice chains with any given sequences with general energy functions. We believe this approach may be useful for computing the free energy landscapes and the thermodynamics of DNA or RNA interactions and the binding to a DNA or RNA target.

ACKNOWLEDGMENTS

S-J. C. thanks Ken A. Dill for useful discussions and encouragement and Francis Schmidt and Om Sehgal for useful discussions, and the University of Missouri Research Board and the American Chemical Society (PRF) for support.

[1] P. Klaff, D. Riesner, and G. Steger, Plant Mol. Bio. **32**, 89 (1996).
[2] J. P. Staley and C. Guthrie, Cell **92**, 315 (1998).
[3] N. Delihas, S. E. Rokita, and P. Zheng, Nat. Biotechnol. **15**, 751 (1997).
[4] C. B. Epsteini and R. A. Butow, Current Opin. Biotechnol. **11**, 36 (2000).
[5] R. M. Knegtel, I. D. Kuntz, and C. M. Oshiro, J. Mol. Biol. **266**, 424 (1997).
[6] D. H. Mathews, M. E. Burkard, S. M. Freier, J. R. Wyatt, and D. H. Turner, RNA **5**, 1458 (1999).
[7] R. Brem and K. A. Dill, Protein Sci. **8**, 1134 (1999).
[8] A. Mattevi, M. Rizzi, and M. Bolognesi, Curr. Opin. Struct. Biol. **6**, 824 (1996).
[9] E. Westhof and D. J. Patel, Curr. Opin. Struct. Biol. **7**, 305 (1997).
[10] P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, J. Mol. Biol. **286**, 593 (1999).
[11] K. A. Dill *et al.*, Protein Sci. **4**, 561 (1995).
[12] S.-J. Chen and K. A. Dill, J. Chem. Phys. **103**, 5802 (1995).
[13] S.-J. Chen and K. A. Dill, J. Chem. Phys. **109**, 4602 (1998).
[14] S.-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **97**, 646 (2000).
[15] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[16] D. C. Poland and H. A. Scheraga, *Theory of the Helix–Coil Transition* (Academic, New York, 1970).
[17] A. S. Benight, R. M. Wartell, and D. K. Howell, Nature (London) **289**, 203 (1981).
[18] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[19] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[20] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[21] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[22] D. W. Miller and K. A. Dill, Protein Sci. **6**, 2166 (1997).
[23] K. A. Dill and H. S. Chan, Nature Struct. Biol. **4**, 10 (1997).
[24] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[25] K. A. Dill, Protein Sci. **8**, 1166 (1999).
[26] R. Owczarzy *et al.*, Biopolymers **44**, 217 (1997).
[27] C. L. Liu, *Introduction to Combinatorial Mathematics* (McGraw–Hill, New York, 1968).
[28] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw–Hill, New York, 1965).
[29] J. S. McCaskill, Biopolymers **29**, 1105 (1990).