Coarse-grained lattice model for investigating the role of cooperativity in molecular recognition

Hans Behringer, Andreas Degenhard, and Friederike Schmid
Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany
(Received 4 July 2007; published 14 September 2007)

Equilibrium aspects of the molecular recognition of rigid biomolecules are investigated using coarse-grained lattice models. The analysis is carried out in two stages. First, an ensemble of probe molecules is designed with respect to the target biomolecule. The recognition ability of the probe ensemble is then investigated by calculating the free energy of association. The influence of cooperative and anticooperative effects accompanying the association of the target and probe molecules is studied. Numerical findings are presented and compared to analytical results which can be obtained in the limit of dominating cooperativity and in the mean-field formulation of the models.

DOI: 10.1103/PhysRevE.76.031914 PACS number(s): 87.15.Aa, 89.20.−a

I. INTRODUCTION

Molecular recognition is the ability of a biomolecule to interact preferentially with a particular target molecule among a vast variety of different but almost identically looking rival molecules. Examples of specific recognition processes comprise enzyme-substrate binding, antibody-antigen binding, protein-receptor interactions, or cell-mediated recognition [1,2]. Molecular recognition is essential for biological systems such as the immune system to work efficiently. Whereas macromolecules are held together by covalent bonds, the recognition process is governed by specific noncovalent interactions such as ionic binding, the van der Waals interaction, the formation of hydrogen bonds, and hydrophobicity [3]. In an aqueous environment those noncovalent bonds contribute an energy of the order of 1–2 kcal/mol with the relatively strong hydrogen bonds sometimes contributing up to 8–10 kcal/mol [4]. The noncovalent bonds are thus only slightly stronger than the thermal energy $k_B T_{\mathrm{room}} \simeq 0.62$ kcal/mol at physiological conditions, and therefore the specificity of biomolecule recognition is only achieved if a large number of functional groups of the two molecules to recognize each other precisely match and thus a large number of noncovalent bonds can be formed [5]. The binding sites of the two molecules are said to be complementary to each other. This view of molecular recognition for inflexible macromolecules is sometimes called the "lock-and-key" mechanism [6]. However, there are notable recognition processes that involve flexible biomolecules [7]. The matching of a large number of functional groups is then achieved by a conformational change giving rise to large entropic contributions (so-called "induced-fit" scheme) [8]. In addition to short-range interactions ensuring the stability of the complex for a sufficiently large time, long-range electrostatic interactions are believed to pre-orient the biomolecules so that the probability of the contact of the complementary patches on the two molecules upon collision is increased [9,10].

An understanding of the principles of recognition processes between biomolecules is not only important from a scientific point of view, but also for biotechnological and biomedical applications. Knowledge of these principles is a necessary input for the design of synthetic heteropolymers with molecular recognition ability so that they can interact with a biological environment—i.e., biomolecules, cells, and tissues—in a programmable way (see, e.g., the review in [7]).

In recent years much effort has been spent investigating the structural basis for the recognition of two rigid proteins [2,9,11–14]. In particular, the recognition sites of the two proteins in contact have been analyzed. The recognition site on a protein basically consists of residues—i.e., amino acids—which interact with residues of the other proteins. It is found to be made up of largely hydrophobic residues so that its hydrophobicity is comparable with that of the interior of the protein. For the development of idealized coarse-grained models, it is therefore assumed that hydrophobicity plays a major role in recognition processes. Consequently, the residue interactions in the idealized models investigated in this article are assumed to be purely of hydrophobic nature.

Investigation of the underlying mechanisms of molecular recognition processes from a physical point of view has recently gained growing interest. In particular, the question of the specificity of recognition processes has been addressed by methods of statistical mechanics [15–26]. Nevertheless, the view of specificity, which is basically the occurrence of a preferential binding of the recognition agents in the presence of a diversity of rival molecules, remains yet incomplete (the introductory remarks in [27] about the diversity of definitions of specificity found in the literature still apply).

In this article we develop coarse-grained lattice models for the investigation of the principles of molecular recognition processes. Our approach, which is described in Sec. II, consists of two stages: In the first step a design of probe molecules is carried out. This step mimics the design in biotechnological applications or the evolution in nature. In the second step the recognition ability is calculated by considering the free energy of association of the probe molecules with the target and a structurally different rival molecule. This general approach is illustrated for a modified hydrophobic-polar (HP) model in Sec. III. In Sec. IV the modified HP model is extended to take cooperative effects on a residue-specific level into account. The resulting model is investigated in its mean-field formulation and in the limiting case of dominant cooperativity, which can be tackled analytically. In addition, the model is investigated numerically for the case where the contributions of the direct residue-residue

interactions and the indirect cooperative interactions are of same order. The findings are compared to the limiting case of dominating cooperativity and to the mean-field findings. In Sec. V another possible way to incorporate cooperative effects is analyzed. The article closes with a conclusion and an outlook (Sec. VI). Among other findings some of the numerical results have been published recently in a separate letter [28].

II. GENERAL APPROACH

In this article we study coarse-grained models for the recognition of two rigid proteins. Under physiological conditions the complex of the proteins is stabilized by noncovalent interactions across its interface. The binding of the proteins is accompanied by a decrease of entropy due to immobilizing translational, rotational, and conformational degrees of freedom. The gain in energy on forming the complex has thus to be strong enough to overcome these entropic costs. In our model the proteins are considered to be rigid so that conformational changes of the backbone of the proteins need not be taken into account. This assumption is fulfilled for a large variety of real protein-protein associations; however, even for the association of two rigid proteins minor rearrangements of the side chains of the amino acids will occur (e.g., [11]).

The energetics at the contact interface of the complex can be formulated in a coarse-grained way where coarse graining is adopted both for the structural properties of the recognition sites of the involved biomolecules and for the interaction between two residues [28]. Consider a recognition site of N residues on both proteins. For simplicity, it is assumed that the two recognition sites contain the same number of residues and that precisely two residues match respectively in the interface. Notice that a recognition site found in natural protein-protein complexes contains typically of the order of 30 residues [9,13]. The chemical structure of the recognition site of the protein to be recognized, which is called the *target* molecule in the following, is characterized in a coarse-grained approach by a discrete variable $\sigma=(\sigma_1,\ldots,\sigma_N)$ where the value of σ_i specifies the type of the residue at positions i, $i=1,\ldots,N$, on the recognition site. Similarly, the types of residues of the recognition site of the other protein which recognizes the target are specified by a second variable $\theta=(\theta_i,\ldots,\theta_N)$. In the following this second biomolecule is called the *probe* molecule. On a coarse-grained level, the interaction of the functional groups across the interface is described by a Hamiltonian $\mathcal{H}(\sigma,\theta;S)$ where we incorporate an additional interaction variable $S=(S_1,\ldots,S_N)$. The variable S_i takes the quality of the contact of the residues of the two proteins at position i into account, where a good contact leads to a favorable contribution to binding and a bad one only to a small contribution. A good contact may imply, for example, that the distance between the two residues is small or the polar moments of residues are appropriately aligned to each other. A steric hindrance, on the other hand, may result in a large distance between the residues and consequently one has a bad contact. The variable S therefore models effects that are related to minor rearrangements of the side-chains of the amino acids when the two proteins form a complex.

Along these general lines a first model—namely, a modification of the HP model—can be formulated. In the HP model only two different types of residues are distinguished—namely, hydrophobic (H) and polar (P) (i.e., hydrophilic) ones—so that the variables σ and θ specify the degree of hydrophobicity of the residues. This restriction to the hydrophobic interaction is motivated by the observation that the hydrophobicity is a major property that discriminates the recognition site from other patches on the surface of a protein. Hydrophobic residues are described by the variable $\sigma_i=+1$ and polar residues by $\sigma_i=-1$. The Hamiltonian is then given by

$$\mathcal{H}_{\mathrm{HP}}(\sigma,\theta;S)=-\varepsilon\sum_{i=1}^{N}\frac{1+S_i}{2}\sigma_i\theta_i, \tag{1}$$

where the sum extents over the N positions of the residues of the recognition site and the interaction constant ε is positive. It is typically of the order of $\varepsilon\simeq 1$ kcal/mol for hydrophobic interactions [3]. The product $-\varepsilon\sigma_i\theta_i$ describes the mutual interaction of the residues in contact across the interface. The additional variable S_i can take on the values ± 1. Thus for $S_i=+1$ one has a good contact, leading to a nonzero contribution to the total interaction energy; for $S_i=-1$, on the other hand, one has a bad contact and no energy contribution. Notice that a good contact does not necessarily lead to a favorable energy contribution. Note also that the original HP model, which has been introduced to study the protein-folding problem [29], does not contain an additional variable S to model the quality of the contact.

The grouping of the 20 natural amino acids into classes of characteristic types is very important for the development of minimal models for the study of protein interactions. The reduction to a hydrophobic and a polar type and thus the use of an Ising-like model Hamiltonian such as (1) on a coarse-grained level is also justified by the findings in [30]. In this work Li *et al.* applied an eigenvalue decomposition to the Miyazawa-Jernigan matrix of the interresidue contact energies of amino acids. They found that the interaction matrix can be parameterized by an Ising-like model where the "spin variable" can take on different discrete values. As these values show, a bimodal distribution the reparametrization basically reduces to the Ising model where the two possible values of the "spins" describe hydrophobic and polar residues. Introducing the additional variable S for the rearrangement of the amino acid side chains, we end up with Hamiltonian (1). Suggested by experimental observations the grouping of the amino acids into five characteristic groups is also widely discussed [31,32]. The reduction in [32], for example, uses a distance-based clustering applied to the Miyazawa-Jernigan matrix. The resulting grouping reproduces the statistical and kinetic features of well-designed sequences in the protein-folding problem. The grouping into five different characteristic types in these approaches points at possible extensions of our model for the contact interaction. In this work, however, we restrict ourselves only to hydrophic and polar residue types.

To study the recognition process between the two biomolecules, we adopt a two-stage approach. First, the structure of the recognition site of the target molecule is fixed to a certain sequence $\sigma^{(0)}=(\sigma_1^{(0)},\ldots,\sigma_N^{(0)})$ of residues. Then this structure is learned by the probe with respect to some learning rules under conditions that are specified by a parameter β_D. This leads to an ensemble of probe molecules of sequences θ at their recognition sites with a probability $P(\theta|\sigma^{(0)})$ depending on the initially fixed target structure. To illustrate this a bit further consider a design step where learning is done just by thermal equilibration. The probability distribution is then technically given by the canonical Boltzmann distribution

$$P(\theta|\sigma^{(0)}) = \frac{1}{Z_D}\sum_S \exp[-\beta_D \mathcal{H}(\sigma^{(0)},\theta;S)], \tag{2}$$

where the normalization Z_D is the usual canonical partition function. The design temperature β_D acts as a Lagrange multiplier that fixes the average energy; however, the parameter β_D may also be interpreted to describe more generally the conditions under which the design has been carried out. This first design step is introduced to mimic the design in biotechnological applications or the process of evolution in nature [33]. Note that in some studies of the protein-folding problem [34,35] and the adsorption of polymers on structured surfaces [36] a similar design step has been incorporated.

In the second step the recognition ability of the designed probe ensemble of structures θ is tested. To this end the ensemble is brought into interaction with both the picked target structure $\sigma^{(0)}$ and a competing (different) structure $\sigma^{(1)}$ at some inverse temperature β which in general is different from the design temperature β_D. The free energy of the probe system interacting with the structure $\sigma^{(\alpha)}$, $\alpha=0,1$, is then

$$F^{(\alpha)} = \sum_\theta F(\theta|\sigma^{(\alpha)})P(\theta|\sigma^{(0)}), \tag{3}$$

where $F(\theta|\sigma^{(\alpha)})$ is the thermal free energy for the interaction between $\sigma^{(\alpha)}$ and a fixed probe sequence θ and an average over the structures in the probe ensemble is carried out. The free energy $F(\theta|\sigma^{(\alpha)})$ is given by

$$F(\theta|\sigma^{(\alpha)}) = -\frac{1}{\beta}\ln\sum_S \exp[-\beta\mathcal{H}(\sigma^{(\alpha)},\theta;S)]. \tag{4}$$

The target with the structure $\sigma^{(0)}$ at its recognition site is recognized by the probe if the associated free energy $F^{(0)}$ is lower than the free energy $F^{(1)}$ for the interaction with the competing structure $\sigma^{(1)}$; i.e., in a mixture of equally many $\sigma^{(0)}$ and $\sigma^{(1)}$ molecules the probe molecules preferentially bind to the original target. This is signaled by a negative free energy difference $\Delta F(\sigma^{(0)},\sigma^{(1)})=F^{(0)}-F^{(1)}$. Thus the specificity of the recognition process is related to the difference between the free energy of association for the competing molecules. For the given structures $\sigma^{(0)}$ and $\sigma^{(1)}$ one can introduce a suitable measure Q for the structural similarity of the target and the rival biomolecule. Carrying out an average over all target and rival structures that are compatible with the specified similarity Q one can compute the averaged free energy difference of association $\Delta F(Q)$ as a function of the similarity between the target and the rival and therefore investigate the overall recognition ability of the model (see Sec. III below for the HP model). Note that in our approach the mechanism which brings the two interacting molecules, in particular the two recognition sites, into contact is not taken into account; i.e., only equilibrium aspects are considered.

In principle, interactions of the residues which do not belong to the recognition sites with solvent molecules have to be considered as well. Solvation effects at the recognition sites and the associated entropy changes are also important for the association process of biomolecules [37,38]. In the coarse-grained model, however, it is assumed that all these contributions are of the same size for all proteins under consideration. Note also that solvation effects are already partially contained in HP models. In addition, the entropic contributions due to a reduction of the translational and rotational degrees of freedom upon forming a complex can be assumed to cancel out in the free energy difference ΔF in a first approximation. This requires at least that the two competing proteins be of comparable shape and size.

In this work we assume that the proteins have the same number of residues at the interface. However, many protein-protein interfaces are curved with different numbers of residues on the two proteins [10]. Nevertheless, we expect our assumption not to be crucial within the above simplified coarse-grained view, at least in a first approximation. As our model characterizes the residues only with respect to their hydrophobicity, one can partition the interface into N contacts and attribute hydrophobicities to the patches on the two proteins that contribute to a particular contact. Then one ends up again with our Hamiltonian (1). For approaches where the residue type is determined by additional features apart form hydrophobicity, correlations between neighboring patches might occur so that our assumption may become questionable.

III. APPLICATION TO A MODIFIED HYDROPHOBIC-POLAR MODEL

The modified HP model of the previous section, can again serve as an illustration of the two-state approach for investigating molecular recognition processes. As (1) does not involve any interaction between neighboring residues of the recognition site of a protein, the two-stage approach can be worked out exactly.

Design by equilibration. For the HP model the design governed by thermal equilibration leads to the conditional probability

$$P(\theta|\sigma^{(0)}) = \frac{\exp\left(\frac{\varepsilon\beta_D}{2}\sum_i \theta_i\sigma_i^{(0)}\right)}{\left[2\cosh\left(\frac{\varepsilon\beta_D}{2}\right)\right]^N} \tag{5}$$

of the structure θ at the recognition site of the probe molecule. As mentioned in the previous section, the design temperature β_D may be interpreted to characterize the conditions under which the design has been carried out. This can be illustrated using the present example of the HP model. In the

HP model the value of σ_i or θ_i, respectively, basically specifies the hydrophobicity of the residue at position i. The total hydrophobicity of the recognition site of the target molecule is then $H^{(0)}=\Sigma_i\sigma_i^{(0)}$. From relation (5) one can calculate the average hydrophobicity $\langle H_D\rangle$ of probe structures:

$$\langle H_D\rangle = \sum_k \sum_\theta \theta_k P(\theta|\sigma^{(0)}) = H^{(0)} \tanh\left(\frac{\varepsilon\beta_D}{2}\right). \tag{6}$$

Thus the Lagrange parameter β_D can be used to fix the average hydrophobicity of the designed probe ensemble, which is achieved by controlling the supply of hydrophobic residues during the design procedure.

The probability distribution (5) for the designed structures θ explicitly depends on the structure $\sigma^{(0)}$ of the recognition site of the fixed target molecule. For the HP model a design under ideal conditions—i.e., $1/\beta_D=0$—the structure θ would simply be a copy of $\sigma^{(0)}$. However, for nonideal conditions with $\beta_D<\infty$ "defects" appear in the design procedure and the obtained structure θ deviates on average from $\sigma^{(0)}$. This deviation can be quantified by the complementarity parameter

$$\mathcal{K}(\theta,\sigma^{(0)}) = \sum_i \theta_i\sigma_i^{(0)}. \tag{7}$$

The possible values of $\mathcal{K}$ range from $-N$ to N in even steps. A value $\mathcal{K}(\theta,\sigma^{(0)})$ close to N means a high complementarity, and the interaction of the probe structure θ with $\sigma^{(0)}$ can lead to a large enough energy decrease so that a complex can be formed. On the other hand, a value of $\mathcal{K}(\theta,\sigma^{(0)})$ much less than N signals a poor match between the two recognition sites and therefore it is unlikely that a complex is stabilized.

The probability distribution $P(\theta|\sigma^{(0)})$ can be converted to a distribution function for the complementarity, leading to the probability

$$P(K) = \sum_\theta P(\theta|\sigma^{(0)})\,\delta_{\mathcal{K}(\theta,\sigma^{(0)}),K} \tag{8}$$

$$= \binom{N}{\frac{1}{2}(N+K)} \frac{\exp\left(\frac{\varepsilon\beta_D}{2}K\right)}{\left[2\cosh\left(\frac{\varepsilon\beta_D}{2}\right)\right]^N} \tag{9}$$

to have a complementarity parameter K in the designed ensemble. The quality of the design can now be measured by the average complementarity of the designed structures θ, which is given by

$$\langle K\rangle = \sum_K KP(K) = N\tanh\left(\frac{\varepsilon\beta_D}{2}\right) \tag{10}$$

for the modified HP model. For large β_D one gets a probe ensemble which is fairly complementary to the fixed target structure. Thus large values of β_D correspond to good design conditions, an observation which can already be deduced from the interpretation of β_D as an inverse temperature. In the hydrophobicity interpretation discussed above large values of β_D signify comparable hydrophobicities of the target and the probe molecule.

Recognition ability. The recognition ability of the probe molecules is tested by comparing the free energy of association with the target structure $\sigma^{(0)}$ and a competing molecule with structure $\sigma^{(1)}$ at its recognition site. For the HP model (1) with its two different types of residues, one can introduce the similarity parameter

$$Q = \sum_i \sigma_i^{(0)}\sigma_i^{(1)}. \tag{11}$$

For Q close to its maximum value N the competing molecule has a recognition site that is almost identical to the one of the target molecule. In terms of the similarity parameter Q the free energy difference is given by

$$\Delta F(Q) = -\frac{1}{2}\varepsilon N\tanh\left(\frac{\varepsilon\beta_D}{2}\right)(N-Q). \tag{12}$$

The free energy difference is always negative as soon as the recognition site of the competing molecule is not identical to the one on the target molecule. In equilibrium, the probe molecule therefore binds preferentially to the target molecule and thus the target molecule is specifically recognized. The difference in free energy increases for an decreasing similarity parameter Q. Note also that the slope of the free energy difference depends only on the conditions under which the design of the probe molecules has been carried out.

IV. ROLE OF COOPERATIVITY IN MOLECULAR RECOGNITION

Cooperative effects play an essential role in many biological processes such as the catalysis of biochemical reactions by enzymes. Cooperativity is presumably also very important for molecular recognition processes [39]. In general, cooperativity means that the binding strength of two residues depends on the binding interactions in the neighborhood of the two residues in contact. Thus the energetic properties of residues when interacting with other residues cannot be inferred by considering them isolated from the local environment. This has implicitly been done, however, in the modified HP model (1) where the interaction constant ε has been attributed to the residue-residue interaction independently of the corresponding local environment.

In this section the modified HP model of the preceding section is extended to incorporate the effect of cooperative interactions on molecular recognition. Note that in Ref. [39] it has been argued that cooperativity should be incorporated on a residue-specific level.

During the association process, rearrangements of the amino acid side chains are observed. Thus in the idealized model applied in this work cooperative effects stem from the behavior of the variables S_i. A possible extension of the modified HP model which takes cooperative interactions into account is given by

$$\mathcal{H}(\sigma,\theta;S) = -\varepsilon\sum_{i=1}^{N}\frac{1+S_i}{2}\sigma_i\theta_i - J\sum_{\langle ij\rangle} S_iS_j. \tag{13}$$

The first sum describes again the hydrophobic interaction, whereas the second sum represents the additional cooperative interaction. It extends over the neighbor positions of the residue at position i. For a fixed i on a square-lattice the sum

includes therefore four terms. The interaction coefficient J is positive for cooperative interactions and negative for anticooperative interactions. To get an impression of its effect consider the design step. Suppose that at position i one has a hydrophobic residue on the target molecule. Then the first term in the HP Hamiltonian (13) favors a hydrophobic residue adsorbing there with a good contact $S_i=+1$ on average. Suppose now that on one of the neighboring positions j of i on the target one has again a hydrophobic residue. If a hydrophobic residue gets adsorbed at the corresponding position of the probe structure, a good contact with $S_j=+1$ is favored by the hydrophobic interaction term in (13). But then the second cooperative term leads to an additional energy decrease for $J>0$. If on the other hand a polar residue shows up at the position j on the probe molecule, the hydrophobic contribution in (13) tries to avoid a contact—i.e., $S_j=-1$, on average—which then leads to an unfavorable energy increase due to the cooperative term. The quality of a contact thus couples to the quality of the contacts in the neighborhood of a residue. For a positive constant J the cooperativity is therefore expected to enhance the fit of the molecules at the interface, resulting in an increased average complementarity compared to an interaction without cooperativity. Similarly, one expects an increase in the recognition specificity. In the subsequent subsections these suggestions are investigated for cooperative interactions.

Note that θ_i (and thus the product $\sigma_i\theta_i$) in Hamiltonian (13) is a random variable whose distribution is obtained by the design step. The energy function (13) describes therefore a random field Ising model. Contrary to the models mostly investigated in the literature (e.g., [40,41]), the distribution function of the random variable $\sigma_i\theta_i$ is not symmetric with respect to a sign reflection.

A. Limiting case of dominant cooperativity

The case where the cooperative contribution to the total energy dominates can be investigated analytically. Consider the situation where $J \gg N\varepsilon$. The cooperative term $-J\Sigma_{\langle ij\rangle}S_iS_j$ in the Hamiltonian (13) has discrete energy levels $-4NJ$, $-4(N-1)J, \ldots, +4NJ$ for a recognition site with a rectangular geometry where each residue has four neighbors. The hydrophobic interaction term $-\varepsilon\Sigma_i\frac{1+S_i}{2}\sigma_i\theta_i$ has also discrete levels ranging from $-N\varepsilon$ to $+N\varepsilon$. For the above assumption $J \gg N\varepsilon$ the global rough structure of the spectrum of the Hamiltonian (13) is basically determined by the cooperative contributions. The hydrophobic interaction of the residues in contact introduces only small variations about the main energy levels with two adjacent ones being separated by an amount of $4J$. For a small temperature—i.e., a large β—the statistical behavior is dominated by the twofold-degenerate lowest-energy state of the cooperative interaction term with all S_i being either in the state +1 or in the state −1. Due to this reduction of the phase space of possible S configurations, the two-stage approach can be worked out analytically. The dominance of the cooperative term leads to the new effective Hamiltonian

$$\mathcal{H} \overset{J\gg\varepsilon N}{\sim} -\frac{1+s}{2}\varepsilon\sum_{i=1}^{N}\sigma_i\theta_i - 4NJ, \tag{14}$$

where the scalar variable s can have the values ±1. The design step now yields the probability distribution

$$P(\theta|\sigma^{(0)}) = \frac{1+\exp\left(\varepsilon\beta_D\sum_i \theta_i\sigma_i^{(0)}\right)}{2^N+[2\cosh(\varepsilon\beta_D)]^N} \tag{15}$$

for the structure of the recognition site of the probe molecules. The corresponding distribution of the complementarity between the structures $\sigma^{(0)}$ and θ is

$$P(K) = \binom{N}{\frac{1}{2}(N+K)}\frac{1+\exp(\varepsilon\beta_D K)}{2^N+[2\cosh(\varepsilon\beta_D)]^N}. \tag{16}$$

The distribution function for the complementarity parameter K can again be used to calculate the average complementarity of the designed molecules. For large N [for which the term 2^N in the denominator of (16) can be neglected as long as $\beta_D \neq 0$] one obtains

$$\langle K\rangle \overset{J\gg\varepsilon N}{\sim} N\tanh(\varepsilon\beta_D). \tag{17}$$

In the situation of a dominating cooperative interaction the average complementarity is increased compared to the case where cooperativity is absent. This suggests that values of the cooperativity interaction constant J comparable to the size of the hydrophobic interaction constant ε might also enhance the quality of the design step. This question is investigated numerically in the subsequent Sec. IV B.

In the second step the designed probe ensemble interacts with the chosen target structure $\sigma^{(0)}$ and a competitive one $\sigma^{(1)}$. The associated free energy averaged with respect to the distribution of the structures θ of the probe molecules is in general given by

$$F^{(\alpha)} = -\frac{1}{\beta}\sum_\theta \ln[1+\exp(\varepsilon\beta\sum_i \theta_i\sigma_i^{(\alpha)})] \tag{18}$$

$$\times\frac{1+\exp(\varepsilon\beta_D\sum_i \theta_i\sigma_i^{(0)})}{2^N+[2\cosh(\varepsilon\beta_D)]^N}. \tag{19}$$

In the case of a large number of residues $N \gg 1$, again further progress can be made analytically. Consider first the free energy of association of the system with the fixed target structure. In this case the sum over the possible structures of the designed probe molecules can be converted into a sum over the complementarity parameter K:

$$F^{(0)} = -\frac{1}{\beta}\sum_K \ln[1+\exp(\varepsilon\beta K)]P(K). \tag{20}$$

The dominant contributions to this sum arise from the values of K close to the maximum of the distribution $P(K)$. For suitably large β_D this maximum, however, occurs for $K \sim O(N)$ and thus it is large as well [compare relation (17)]. Therefore, in the limit $N \gg 1$ one can use the replacements

$1+\exp(\beta_D \varepsilon K) \approx \exp(\beta_D \varepsilon K)$ and $\ln[1+\exp(\beta \varepsilon K)] \approx \beta \varepsilon K$. Using these approximations the summation in (20) leads to

$$F^{(0)} \overset{N\gg 1}{\sim} -\varepsilon \sum_K K P(K) = -\varepsilon \langle K \rangle \overset{(17)}{=} -\varepsilon N \tanh(\varepsilon \beta_D). \tag{21}$$

A similar conversion cannot be applied to the summation over the designed molecules in the calculation of $F^{(1)}$ as both $\theta_i \sigma_i^{(0)}$ and $\theta_i \sigma_i^{(1)}$ terms appear. Defining the auxiliary variables $k_i := \theta_i \sigma_i^{(0)}$ and $q_i := \sigma_i^{(0)} \sigma_i^{(1)}$ and noting that $(\sigma_i^{(\alpha)})^2 = 1$, the free energy $F^{(1)}$ is explicitly given by

$$F^{(1)} = -\frac{1}{\beta} \sum_k \ln[1 + \exp(\varepsilon \beta \sum_i k_i q_i)] \tag{22}$$

$$\times \frac{1 + \exp(\varepsilon \beta_D \sum_i k_i)}{2^N + [2 \cosh(\varepsilon \beta_D)]^N}. \tag{23}$$

The variable k_i specifies the local complementarity between the target $\sigma^{(0)}$ and a particular probe structure θ. Using again the observation that the dominant contributions originate from values of large $K = \Sigma_i k_i$, one can use again the replacement $1+\exp(\beta_D \varepsilon K) \approx \exp(\beta_D \varepsilon K)$. The logarithmic factor in (22) gives large contributions if the majority of the q_i variables are in state +1. Thus, in the limit of $Q = \Sigma_i q_i \gg 1$ the sum in (22) can be worked out and one obtains

$$F^{(1)} \overset{N,Q\gg 1}{\sim} -\varepsilon Q \tanh(\varepsilon \beta_D). \tag{24}$$

The free energy difference in terms of the similarity Q of the competing molecules $\sigma^{(0)}$ and $\sigma^{(1)}$ is now given by

$$\Delta F \overset{N,Q\gg 1}{\sim} -\varepsilon \tanh(\varepsilon \beta_D)(N - Q) \tag{25}$$

for positive and large Q. Again, one has a linear dependence in the vicinity of $Q=N$. This can be compared to the corresponding result (12) for the situation with $J=0$. The cooperativity increases the slope of the free energy difference, and thus the recognition ability of the designed probe ensemble is increased by cooperativity.

In the limit $Q = \Sigma_i q_i \ll -1$, on the other hand, almost all q_i take on the value -1 and thus $\Sigma_i k_i q_i$ is close to $-N$ for those k_i leading to the dominant contributions in (22). One therefore has

$$\ln[1 + \exp(\varepsilon \beta \sum_i k_i q_i)] \overset{N\gg 1, Q\ll -1}{\sim} \exp(-\varepsilon \beta N) \tag{26}$$

for the logarithmic factor of the dominant terms in (22). The free energy of association of the probe molecules with the rival molecule is thus $F^{(1)} \sim O(e^{-N})$ so that

$$\Delta F \overset{N\gg 1, Q\ll -1}{\sim} F^{(0)} = -\varepsilon N \tanh(\varepsilon \beta_D). \tag{27}$$

In the limit $Q \ll -1$ the free energy difference is thus independent of the similarity parameter Q between the target structure and the rival structure.

For a similarity parameter $|Q| \sim O(1)$ one expects deviations form the behavior for large $|Q|$. For the free energy difference per residue, $\Delta F/N$, as a function of the similarity per residue, Q/N, however, the deviations show up for similarities Q/N of the order of $1/N$. The free energy difference per residue will thus develop a kink at $Q/N=0$ in the asymptotic limit of $N \to \infty$ so that it is given by expression (25) for positive Q/N and by relation (27) for negative Q/N. The range of values of the similarity per residue where deviations between the free energy for a system with finite N and the asymptotic result show up is shrinking for increasing N.

B. Numerical results for arbitrary cooperativity

The above analysis of the limiting case $J \gg N\varepsilon$ with a dominant cooperative interaction suggests that cooperativity enhances the quality of the design step and eventually increases the recognition ability. In this section this suggestion is investigated more closely for cooperativity constants J which are of the order of the interaction constant ε of the hydrophobic interaction term in (13).

Design. For nonzero, but finite values of J it is not possible any more to solve the model analytically. Therefore, the two-stage approach has to be carried out numerically. To this end the density of states for the design step is calculated as a function of the energy and the complementarity parameter for a fixed cooperativity J. The density of states is generally given by

$$\Omega_J(K;E) = \sum_{\theta,s} \delta_{K,\mathcal{K}(\theta,\sigma^{(0)})} \delta_{E,\mathcal{H}(\theta,\sigma^{(0)};S)} \tag{28}$$

for a fixed target structure $\sigma^{(0)}$. The density of states, $\Omega_J(K;E)$, is thus the number of (θ,S) configurations that have energy E when interacting with the target and a complementarity K of the probe molecule θ to the target recognition site. In general, the density of states depends additionally on the configuration $\sigma^{(0)}$ of the recognition site of the target molecule. However, for the HP model (13) the density of states has no explicit dependence on $\sigma^{(0)}$ as the variables θ_i can be transformed to the auxiliary variables $k_i := \sigma_i^{(0)} \theta_i$, which have the same phase space as θ_i, so that $\sigma^{(0)}$ does not appear anymore.

The density of states can be calculated directly using efficient Monte Carlo algorithms [42–44]. In this work the Wang-Landau algorithm has been applied. Once the density of states is known, the probability distribution of the complementarity is basically obtained by calculating the Laplace transform of Ω_J, giving, up to a normalization,

$$P_J(K;\beta_D) \sim \sum_E \Omega_J(K,E) \exp(-\beta_D E). \tag{29}$$

From this distribution function one can calculate the average complementarity $\langle K \rangle(J) = \Sigma_K P_J(K;\beta_D) K$, which is shown in Fig. 1. The calculations have been carried out for a square-lattice geometry with $N=256$ residues. We have checked that the curves show only minor finite-size effects for recognition sites of realistic sizes with $N \sim O(30)$ (see [28]). The qualitative findings discussed in the following are independent of the number N of residues involved in the interface. It can be seen that cooperativity increases the average complementar-

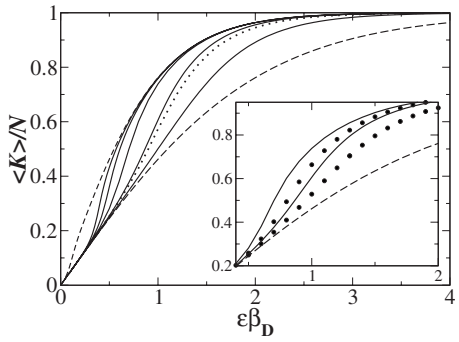

FIG. 1. Average complementarity per site of the designed ensemble for the HP model (13) for different values of J. The lower dashed curve corresponds to $J=0$; the upper dashed line represents the limiting case $J\to\infty$ (for large N). The solid curves in between from the bottom up correspond to the values 0.1, 0.25, 0.5, 0.75, and 1.0 of J in units of ε. The dotted curve shows the result for a system with additional next-nearest-neighbor cooperativity with $J_{\mathrm{NNN}}/\varepsilon=J_{\mathrm{NN}}/\varepsilon=0.1$. The inset compares the numerical results (solid lines) with the mean-field findings of Sec. IV C (circles) for cooperativities 0.25 and 0.5 from the bottom up. The dashed curve corresponds again to $J=0$.

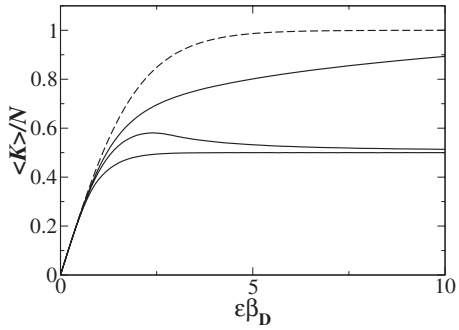

FIG. 2. Average complementarity of the probe ensemble for the anticooperative HP model (13) with $J<0$. The dashed line represents $J=0$; the solid curves from top to bottom correspond to the values −0.1, −0.2, and −0.5 of J in units of ε.

ity of the probe molecules for large enough values of β_{D}. For a parameter value of the order of $\varepsilon\beta_{\mathrm{D}}\simeq 1$, a small change in the cooperativity J leads to a large difference in the average complementarity. Therefore, small changes in J can have a large impact on the quality of the design step. As the typical energy ε of a noncovalent bond is of the order of 1 kcal/mol, this regime corresponds indeed to physiological conditions.

The Hamiltonian (13) contains a cooperative term where the quality of the contact couples to the contact variable at the neighboring sites. This limitation to nearest-neighbor interactions can be relaxed by allowing additional couplings to sites that are farther away. As long as the range of the cooperative coupling is finite, however, we expect that the average complementarity $\langle K\rangle$ is qualitatively similar as for model (13). For the system with nearest- and next-nearest-neighbor interactions (with the same constant J), the case of dominant cooperativity can be treated as above (Sec. IV A), leading to the same effective Hamiltonian (14) with the irrelevant constant replaced by $-8NJ$. So the same limiting curves for $\langle K\rangle$ as well as ΔF result. However, the additional interactions have the consequence that the maximum effect of cooperativity will already show up for smaller values of J. This is shown in Fig. 1 for the model with additional next-nearest-neighbor cooperativity.

Before analyzing the recognition ability for $J\neq 0$, consider briefly the influence of an anticooperative interaction with $J<0$ in Hamiltonian (13) on the average complementarity $\langle K\rangle$. From the discussion of the effect of the cooperative term within the design step, one may expect that anticooperative interactions should decrease $\langle K\rangle$. For a probe molecule with a high complementarity to the target molecule, all S_i tend to be in state +1 to ensure good contacts and thus a large energy decrease due to the hydrophobic interaction. However, the anticooperative term then leads to an energy increase so that the two contributions to the Hamiltonian (13) compete with each other. Two different regimes can now be expected. Large values of the parameter β_{D} favor structures θ that are highly complementary to the target $\sigma^{(0)}$. For $0>J>-\varepsilon/8$ the hydrophobic interaction term is dominant, leading to a majority of good contacts $S_i=+1$, and thus $\langle K\rangle$ is expected to become N for increasing β_{D}. However, if $J<-\varepsilon/8$, the second anticooperative term dominates, leading to an alternating structure of good and bad contacts where the S_i of two neighboring positions have different signs. Note that in such a situation the direct hydrophobic-polar interaction contributes a maximum favorable energy $-\varepsilon/2$ per site, whereas the cooperative term gives the maximum contribution $4J$ per site giving the crossover value $J=-\varepsilon/8$ for the considered square geometry. For one-half of the residues one therefore has preferably good contacts so that the residue on the probe molecule is of the same type as the one on the target molecule on average. For the other half of the positions, however, one has bad contacts. For those positions the hydrophobic interaction term in (13) does not contribute and the probabilities of the residue on θ to be hydrophobic or polar at such positions are equal. For $J<-\varepsilon/8$ one thus expects that $\langle K\rangle$ tends to $N/2$ for increasing β_{D}. These expectations are indeed confirmed by numerical investigations as shown in Fig. 2.

In the general discussion of the extended model (13), it has been argued that the cooperative term will increase the effective contribution of a residue-residue contact at the interface between the two biomolecules. To get an impression of this increase one can define an effective residue-residue interaction constant $\varepsilon_{\mathrm{eff}}(\beta_{\mathrm{D}},J):=\langle\mathcal{H}(J)\rangle/\langle\mathcal{H}_{\mathrm{HP}}\rangle$ by considering the average interaction energy of the probe ensemble with the target molecule for different values of the cooperativity J. Figure 3 shows that this effective interaction constant is indeed increased by the cooperative term in the Hamiltonian (13).

Recognition ability. Knowledge of the density of states allows the calculation of the recognition ability quantified by the free energy difference

$$\Delta F(Q)=\frac{\langle\delta_{Q,\Sigma_i\sigma_i^{(0)}\sigma_i^{(1)}}\Delta F(\boldsymbol{\sigma}^{(0)},\boldsymbol{\sigma}^{(1)})\rangle_{\sigma^{(0)},\sigma^{(1)}}}{\langle\delta_{Q,\Sigma_i\sigma_i^{(0)}\sigma_i^{(1)}}\rangle_{\sigma^{(0)},\sigma^{(1)}}} \tag{30}$$

for the association of probe molecules with the two structures $\sigma^{(0)}$ and $\sigma^{(1)}$. The results are shown in Fig. 4 for dif-

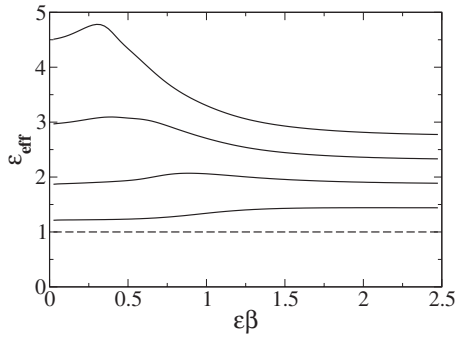

FIG. 3. Effective interaction constant ε_{eff} as defined in the main text as a function of the temperature for the model (13). The solid curves correspond to values 0.25, 0.5, 0.75, and 1.0 of J (in units of ε) from the bottom up.

ferent values of the J. For comparison, the free energy difference for the system with additional next-nearest-neighbor cooperativity is shown as well. An increase in J increases the free energy difference and therefore the recognition specificity of the probe molecules. For a value of J of the order ε the maximum effect of cooperativity has already been reached for the considered temperature values $\beta_D=\beta=1.0$. Thus, the expected increase of the recognition ability by cooperativity for constants $J\simeq\varepsilon$ is indeed confirmed by the numerical results.

To study the influence of different cooperativities on the recognition ability in a more direct way the following approach can be adopted. The cooperativity already influences the design step and optimizes the probe ensemble with respect to the original target structure as can be seen by the dependence of $\langle K\rangle$ on the J. This better optimization influences the testing step as well. In order to investigate the pure influence of the cooperative interaction on the recognition ability more closely, one can use probe ensembles where the average complementarity is fixed to some values K_0 for different values of the J. For comparison, the free energy difference for the system with additional next-nearest-neighbor cooperativity is shown as well. An increase in J increases the free energy difference and therefore the recognition specificity of the probe molecules. For a value of J of the order ε the maximum effect of cooperativity has already been reached for the considered temperature values $\beta_D=\beta=1.0$. Thus, the expected increase of the recognition ability by cooperativity for constants $J\simeq\varepsilon$ is indeed confirmed by the numerical results.

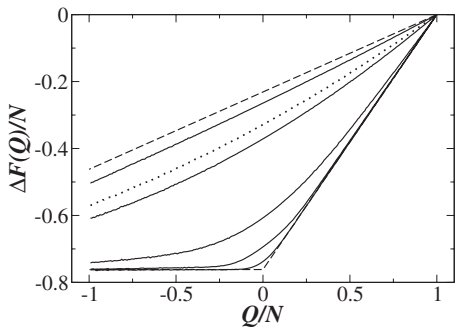

FIG. 4. Free energy difference per site of the association of the probe ensemble with the two competing molecules as a function of their similarity for different cooperativities J in (13). The upper dashed line corresponds to $J=0$. The lower dashed line describes the limiting case $J\to\infty$ in the limit of large N (Sec. IV A). The solid curves from top to bottom correspond to the same values of J as in Fig. 1. The dotted curve shows the result for a system with additional next-nearest-neighbor cooperativity with $J_{\text{NNN}}/\varepsilon=J_{\text{NN}}/\varepsilon=0.1$. The parameters β_D and β are both 1.0.

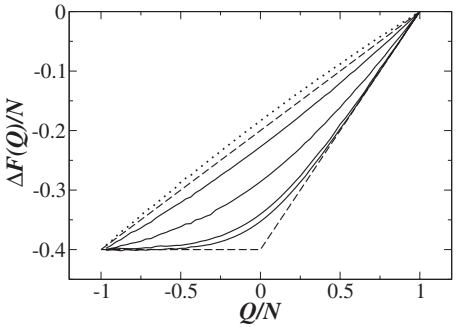

FIG. 5. Free energy difference as a function of the similarity for different cooperativities J (with $\beta=1.0$) where the probe ensemble has been designed to have a fixed $\langle K\rangle/N=0.4$. The upper dashed lines corresponds to $J=0$; the lower one describes the limiting case $J\to\infty$ (and large N). The values of J/ε in (13) are 0.25, 0.5, 0.75, and 1.0 for the solid curves from top to bottom. For the dotted line $J/\varepsilon=-1/2$.

ferent J. This can be done by carrying out the design of the probe molecules at different design temperatures such that $\langle K\rangle(\beta_D,J)=K_0$. The probability distributions obtained when this additional constraint is applied are then used to calculated the difference of the free energy of association of the probe molecules with both the target and the rival molecule. The results are shown in Fig. 5 for recognition sites with $N=64$ residues. Again, it can be seen that an increase in the cooperativity J increases the free energy difference for a fixed similarity Q/N between the target and the rival biomolecule. The dashed lines in Fig. 5 represent the free energy difference for $J=0$ and for the asymptotic regime $J\to\infty$ with $N\gg1$. For large Q/N and large J the free energy difference is already well represented by the asymptotic result. For a cooperativity $J\simeq\varepsilon$ the maximum effect is thus already achieved.

For the minimum similarity parameter $Q=-N$ the free energy difference at fixed K_0 is independent of the cooperativity J (compare Fig. 5). To see this consider the fixed structure $\sigma^{(0)}$ of the recognition site of the target. As the similarity parameter Q is minimum, the competing molecule has the structure $\sigma^{(1)}=-\sigma^{(0)}$ at its recognition site. The free energy difference of association is then given by $\Delta F(-N)=-\frac{1}{\beta}\Sigma_\theta P_J(\theta|\sigma^{(0)})[\ln Z(\theta|\sigma^{(0)})-\ln Z(\theta|-\sigma^{(0)})]$. The partition function $Z(\theta|\sigma^{(1)})=Z(\theta|\sigma^{(0)})$ related to the rival structure explicitly reads

$$Z(\theta|\sigma^{(0)})=\exp\left(-\frac{\beta\varepsilon}{2}\sum_i\sigma_i^{(0)}\theta_i\right) \tag{31}$$

$$\times\sum_S\exp\left(-\beta\varepsilon\sum_i\frac{S_i}{2}\sigma_i^{(0)}\theta_i+\beta J\sum_{\langle ij\rangle}S_iS_j\right) \tag{32}$$

$$=\exp(-\beta\varepsilon\sum_i\sigma_i^{(0)}\theta_i)Z(\theta|\sigma^{(0)}), \tag{33}$$

where a transformation $S_i\to-\tilde{S}_i$ has been used for the last equality. Note that the phase space for $\tilde{S}$ is the same as for

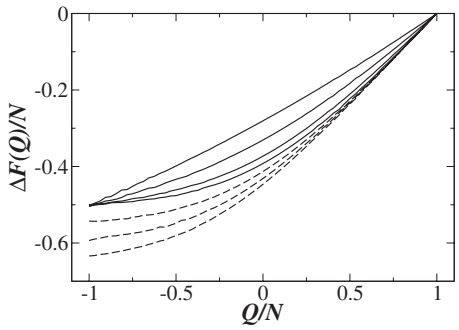

FIG. 6. Free energy difference as a function of the similarity for fixed $J/\varepsilon=1/2$ in (13) and different testing temperatures β. The probe ensemble has been designed to have $\langle K\rangle/N=0.5$. The solid curves correspond to the $\beta=0.5$, 0.75, 1.0, and 1.25 from top to bottom. For the dashed curves $\beta=1.25$ and an additional field γ has been applied—namely, from top to bottom, $\gamma/\varepsilon=0.05$, 0.1, and 0.15.

the variable S. Thus the free energy difference at $Q=-N$ is generally given by

$$\Delta F(Q=-N)=-\varepsilon\sum_K P_J(K;\beta_D)K=-\varepsilon\langle K\rangle(J). \quad (34)$$

As the average $\langle K\rangle$ is fixed to the value K_0 for different J, the free energy difference is the same for all J.

For the HP model with pure hydrophobic interactions the free energy difference is independent of the conditions under which the recognition ability is tested. It is only determined by the design conditions [compare relation (12)]. For the extended HP model (13) with cooperative interactions, this is no longer the case. Apart from the design conditions, encoded in the Lagrange parameter β_D, the free energy difference depends on β, which specifies the conditions for the testing step. In Fig. 6 the free energy difference is shown for different values of β. The cooperativity constant is fixed to be $J/\varepsilon=1/2$; the design temperature β_D is chosen to have $\langle K\rangle=N/2$. For increasing parameters β the absolute value of the free energy difference is increased. For the minimum similarity $Q=-N$ the free energy difference becomes independent of β. Its value at the minimum similarity is only determined by the design conditions and is given by $\Delta F(Q=-N)=-\varepsilon\langle K\rangle$ as has been shown above.

The independence of $\Delta F(Q=-N)$ of the testing temperature β is a result of the symmetry of the underlying model (13). This symmetry is broken by introducing a fieldlike term $-\Sigma_i\gamma S_i$ to the energy. It is expected that there is some bias toward good or bad contacts, leading to a such an additional field with $\gamma\neq 0$. For positive fields γ the recognition ability is again expected to be increased with respect to the situation where γ vanishes. This is shown by the dotted lines in Fig. 6.

C. Mean-field theory for arbitrary cooperativity

After having analyzed the influence of the cooperative terms in the previous sections by means of an asymptotic analysis and Monte Carlo simulations, we briefly sketch how a mean-field treatment can be carried out [45]. The discussion will be restricted to the determination of the averaged complementarity. As already mentioned, the variable $\sigma_i\theta_i$ acts as a random field in (13) and therefore techniques from the theory of disordered systems have to be applied in the mean-field treatment (see, for example, [40,41]). Thus the auxiliary variable $k_i=\theta_i\sigma_i^{(0)}$ which has been introduced in Sec. IV A and specifies a complementarity configuration $k=(k_1,\ldots,k_N)$ is used in the following. The mean-field approach consists of two steps—namely, an equivalent neighbor approximation of the cooperative interaction term and an asymptotic evaluation of the partition sum for large N. The equivalent neighbor approximation of the Hamiltonian (13) reads

$$\mathcal{H}_{EN}=-\frac{J}{2N}\left(\sum_i S_i\right)^2-\varepsilon\sum_i\frac{1+S_i}{2}k_i. \quad (35)$$

We aim at a calculation of the averaged complementarity $K=\langle\Sigma_i k_i\rangle$ containing a thermal average with respect to the interaction variable S and an average over the possible complementarity configurations k of the probe molecules with respect to the target. The thermal average leads to the distribution $P(\theta|\sigma^{(0)})$ of probe molecules and thus to a distribution $P(k)$ of the complementarity configuration itself. Consider first the thermal average with respect to S. The variable $x:=\Sigma_i S_i$ appears quadratically in (35). By introducing an additional auxiliary variable y, it can be linearized in the argument of the Boltzmann factor in the partition sum $Z(k)=\Sigma_S\exp(-\beta\mathcal{H}_{EN})$ with the help of the identity

$$\exp\left(\frac{a}{2N}x^2\right)=\int_{-\infty}^{+\infty}dy\sqrt{\frac{Na}{2\pi}}\exp\left(-\frac{Na}{2}y^2+axy\right), \quad (36)$$

often called the Hubbard-Stratonovich transformation in the literature. Note that the distribution function $P(k)$ of the complementarity configuration is determined by $Z(k)$ up to the normalization. The summation over S can then be carried out, leading to

$$Z(k)\sim\exp\left(\frac{\beta\varepsilon}{2}\sum_i k_i\right)\int_{-\infty}^{+\infty}dy\,\exp[\mathcal{A}(y,k)], \quad (37)$$

with

$$\mathcal{A}(y,k)=-\frac{\beta JN}{2}y^2+\sum_i\ln\cosh\left(\beta Jy+\frac{\beta\varepsilon}{2}k_i\right). \quad (38)$$

In the asymptotic limit of large N the integration over the auxiliary field y in (37) can be carried out using the Laplace method (e.g., [46,47]). This gives

$$Z(k)\sim\exp\left(\frac{\beta\varepsilon}{2}\sum_i k_i+\mathcal{A}(y_0,k)\right) \quad (39)$$

aside from irrelevant factors. The mean field y_0 is determined by the saddle point equation

$$y_0 = \frac{1}{N}\sum_i \tanh\left(\beta J y_0 + \frac{\beta\varepsilon}{2}k_i\right). \tag{40}$$

Note that the mean field explicitly depends on the local complementarity configuration k. These two equations can be used to carry out the configurational average over all k to obtain the averaged complementarity $\langle K\rangle$ by noting that a particular configuration k contains $K^{(+)}$ sites with $k_i=+1$ and $K^{(-)}$ ones with $k_i=-1$. The partition function Z (and thus the distribution function P) as well as the mean field y_0 are therefore only functions of $K^{(\pm)}$. The mean field $y_0(K^{(+)},K^{(-)})$, for example, is then given by

$$y_0 = \frac{K^{(+)}}{N}\tanh\left(\beta J y_0 + \frac{\beta\varepsilon}{2}\right) + \frac{K^{(-)}}{N}\tanh\left(\beta J y_0 - \frac{\beta\varepsilon}{2}\right). \tag{41}$$

The average over k can thus be converted to an average over $(K^{(+)},K^{(-)})$ so that the complementarity $\langle K\rangle=\langle K^{(+)}-K^{(-)}\rangle$ can by worked out using a computer algebra program. The results are shown in the inset of Fig. 1 together with the Monte Carlo findings discussed in the previous paragraph. The mean-field curves behave qualitatively similarly as the Monte Carlo curves. Using a similar decomposition of the similarity configuration $q_i=\sigma_i^{(0)}\sigma_i^{(1)}$ between the target and the rival structure into positive contributions $Q^{(+)}$ and negative ones $Q^{(-)}$, one can calculate the averaged free energy difference $\Delta F(Q)$ [compare relation (30)]. The resulting curves show again the same qualitative behavior as the results from the Monte Carlo simulations.

V. COOPERATIVITY COUPLING TO RESIDUE STRUCTURE

The importance of cooperativity in biological situations was emphasized at the beginning of Sec. IV. In Hamiltonian (13) an additional cooperative term has been introduced, which, however, does not couple to the residue distributions on the recognition sites of the two molecules in contact. In general, the additional cooperative interaction terms might also couple to the structures σ and θ of the target and probe molecules, respectively. One possible coupling is given by the Hamiltonian

$$\mathcal{H}(\sigma,\theta;S) = -\sum_{i=1}^{N}\left(\varepsilon\frac{1+S_i}{2} + J\sum_{i_\delta} S_i S_{i_\delta}\right)\sigma_i\theta_i. \tag{42}$$

The sum in the second term extends over the neighboring positions i_δ of position i on the interface. Again, the cooperative term will lead to an additional energy contribution depending on how the side chains are rearranged in the interface. In the case of a favorable direct energy contribution from the hydrophobic interaction at site i described by the product $\sigma_i\theta_i$, the cooperative term rewards good contacts like in Hamiltonian (13). However, in (13) two neighboring bad contacts due to an unfavorable hydrophobic-polar interaction are also attributed a favorable cooperative contribution. This is no longer the case in Hamiltonian (42) as the sign of the cooperative energy contribution now depends on the sign of the hydrophobic interaction energy at position i on the interface. It is thus expected that the cooperative terms in (42) lead to a more favorable cooperative contribution than those in Hamiltonian (13). The cooperative terms in Hamiltonians (42) are only two possible ways to take into account cooperativity, corresponding in our modeling to mutual interaction of neighboring variables S_i; other extensions are possible as well.

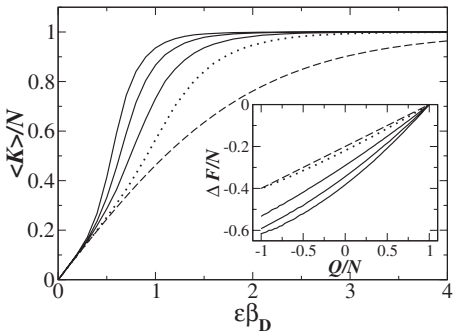

FIG. 7. Averaged complementarity of the probe ensemble designed according to the model (42). The dashed curve represents $J=0$, for the solid curves $J/\varepsilon=0.2$, 0.3, and 0.4 from bottom up. For comparison, the dotted line depicts the corresponding curve for the model (13) with $J/\varepsilon=0.2$. The inset shows free energy differences as a function of the similarity Q for the same set of parameters, where β_D is chosen to have $\langle K\rangle/N=0.4$ (with $\beta=1.0$ for each curve).

As already remarked, the variable $\sigma_i\theta_i$ in (13) is basically a random field the distribution of which is determined by the design step. The model (13) is thus a random field Ising model where the random field $\sigma_i\theta_i$ is asymmetrically distributed. In Hamiltonian (42) this random variable now also couples to the exchange constant J of the interactions between neighboring variables S_i and thus the exchange constant also becomes a random variable. The model (42) is thus an Edward-Anderson-like model in a random field with an asymmetrically distributed exchange constant $J\sigma_i\theta_i$.

The two-stage approach to obtain the recognition ability can now be carried out numerically for the model (42) by calculating again density of states, $\Omega_J(K;E)$, by a Monte Carlo simulation. The results for the averaged complementarity of the probe molecules and the free energy difference are depicted in Fig. 7. One observes a similar qualitative behavior as the corresponding curves for the model (13). Again, it is found that an increase of the parameter J increases the quality of the design step in the sense that the probe molecules are better optimized with respect to the target biomolecule indicated by an increase of $\langle K\rangle$ for higher values of J. Similarly, the recognition ability measured by the free energy difference $\Delta F=F^{(0)}-F^{(1)}$ for a given similarity Q between the target and the rival grows for increasing J.

VI. CONCLUSIONS AND OUTLOOK

We have presented coarse-grained models to study the properties of molecular recognition processes between rigid biomolecules. The development of the models has been mo-

tivated by experimental investigations on the biochemical structure of the interface of protein complexes. A two-stage approach containing a design of probe molecules and a testing of their recognition ability has been adopted. This approach has been used to investigate the role of cooperativity in molecular recognition. The coarse-grained models capture the effects of cooperativity on a residue-specific level. The necessity of such an approach has been pointed out in the literature [39]. We have shown numerical results and compared them to analytic results obtained in the asymptotic limit where cooperative interactions dominate over direct hydrophobic interactions between the residues and in the mean-field formulation of the models. It turned out that a small contribution due to cooperativity can already substantially influence the recognition ability, corroborating the suggestion that cooperativity has a considerable effect on the recognition specificity. Two possibilities to include cooperative interactions have been explicitly analyzed, leading to similar qualitative results. We note in passing that the proposed coarse-grained model can reproduce qualitatively the experimental observation that in antigen-antibody complexes, which require a relatively high binding flexibility, a small number of strong noncovalent bonds across the interface seems to be favored compared to a situation with many but rather weak bonds. The details are published elsewhere [28].

The proposed approach to study molecular recognition with coarse-grained lattice models can be extended in various ways. Apart from working with refined models, which capture more details of the actual physical interactions across the interface of the two biomolecules, the design step can be modified to mimic natural evolution in a more realistic manner. The analysis presented considered on the level of the target and the rival molecule is basically a single-molecule approach, although the molecules are described in a very-coarse-grained way. The influence of the heterogeneity of the mixture of target and rival molecules encountered in real physiological situations as found in a cell, for example, can be incorporated in our analysis. To this end ensembles of targets and rivals differing in certain properties, such as, for example, correlations and length scales, have to be considered. A recent study indeed indicates that the local small-scale structure related to the distribution of the hydrophobicity on the recognition site of the biomolecules seems to play a crucial role in molecular recognition [22].

ACKNOWLEDGMENT

We thank the Deutsche Forschungsgemeinschaft (Grant No. SFB 613) for financial support.

[1] B. Alberts, D. Bray, L. Lewis, M. Raf, K. Roberts, and J. Watson, *Molecular Biology of the Cell* (Garland Publishing, New York, 1994).
[2] *Protein-Protein Recognition*, edited by C. Kleanthous (Oxford University Press, Oxford, 2000).
[3] K. Sneppen and G. Zocchi, *Physics in Molecular Biology* (Cambridge University Press, Cambridge, England, 2005).
[4] M. Delaage, in *Molecular Recognition Mechanisms*, edited by M. Delaage (VCH, New York, 1991), p. 1.
[5] L. Pauling and M. Delbrück, Science **92**, 77 (1940).
[6] E. Fischer, Ber. Dtsch. Chem. Ges. **27**, 2984 (1894).
[7] N. A. Peppas and Y. Huang, Pharm. Res. **19**, 578 (2002).
[8] D. E. Koshland, Proc. Natl. Acad. Sci. U.S.A. **44**, 98 (1958).
[9] J. Janin, in *Protein-Protein Recognition*, edited by C. Kleanthous (Oxford University Press, Oxford, 2000), p. 1.
[10] S. Wodak and S. J. Janin, Adv. Protein Chem. **61**, 9 (2003).
[11] J. Janin and C. Chothia, J. Biol. Chem. **265**, 16027 (1990).
[12] S. Jones and J. M. Thornton, Proc. Natl. Acad. Sci. U.S.A. **93**, 13 (1996).
[13] S. Jones and J. M. Thornton, in *Protein-Protein Recognition*, edited by C. Kleanthous (Oxford University Press, Oxford, 2000), p. 33.
[14] P. Chakrabarti and J. Janin, Proteins Struct. Funct. Genet. **47**, 334 (2002).
[15] D. Lancet, E. Sadovsky, and E. Seidemann, Proc. Natl. Acad. Sci. U.S.A. **90**, 3715 (1993).
[16] J. Janin, Proteins Struct. Funct. Genet. **25**, 438 (1996).
[17] J. Janin, Proteins Struct. Funct. Genet. **28**, 153 (1997).
[18] S. Rosenwald, R. Kafri, and D. Lancet, J. Theor. Biol. **216**, 327 (2002).
[19] J. Wang and G. M. Verkhivker, Phys. Rev. Lett. **90**, 188101 (2003).
[20] A. Polotsky, A. Degenhard, and F. Schmid, J. Chem. Phys. **120**, 6246 (2004).
[21] A. Polotsky, A. Degenhard, and F. Schmid, J. Chem. Phys. **121**, 4853 (2004).
[22] T. Bogner, A. Degenhard, and F. Schmid, Phys. Rev. Lett. **93**, 268108 (2004).
[23] J. Bernauer, A. Poupon, J. Azé, and J. Janin, Phys. Biol. **2**, S17 (2005).
[24] J. Wang, Q. Lu, and H. P. Lu, PLOS Comput. Biol. **2**, e78 (2006).
[25] H. Behringer, T. Bogner, A. Polotsky, A. Degenhard, and F. Schmid, J. Biotechnol. **129**, 268 (2007).
[26] E. Baake, F. den Hollander, and N. Zint, http://arxiv.org/abs/q-bio/0605016.
[27] P. H. Von Hippel and O. G. Berg, Proc. Natl. Acad. Sci. U.S.A. **83**, 1608 (1986).
[28] H. Behringer, A. Degenhard, and F. Schmid, Phys. Rev. Lett. **97**, 128101 (2006).
[29] K. A. Dill, Biochemistry **24**, 1501 (1985).
[30] H. Li, C. Tang, and N. S. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).
[31] J. Wang and W. Wang, Nat. Struct. Biol. **6**, 1033 (1999).
[32] M. Cieplak, N. S. Holter, A. Maritan, and J. R. Banavar, J. Chem. Phys. **114**, 1420 (2001).
[33] In evolution models, however, where sequences reproduce according to the modified HP Hamiltonian (1) and mutate subject to a reversible Markov process, the equilibrium distribution is indeed the Boltzmann distribution [E. Baake and N. Zint (private communication)].

[34] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Biophys. J. **73**, 3192 (1997).
[35] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).
[36] A. Jayaraman, C. K. Hall, and J. Genzer, Phys. Rev. Lett. **94**, 078103 (2005).
[37] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. **72**, 1047 (1997).
[38] M. B. Jackson, *Molecular and Cellular Biophysics* (Cambridge University Press, Cambridge, England, 2006).
[39] E. di Cera, Chem. Rev. **98**, 1563 (1998).
[40] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems* (Cambridge University Press, Cambridge, England, 2001).
[41] T. Schneider and E. Pytte, Phys. Rev. B **15**, 1519 (1977).
[42] A. Hüller and M. Pleimling, Int. J. Mod. Phys. C **13**, 947 (2002).
[43] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[44] D. P. Landau, S. Tsai, and M. Exler, Am. J. Phys. **72**, 1294 (2004).
[45] We are currently investigating further physical aspects of the introduced coarse-grained models for molecular recognition with mean-field methods. More details of the mean-field calculations will be published in a corresponding article.
[46] K. Jänich, *Analysis für Physiker und Ingenieure* (Springer-Verlag, Berlin, 2001).
[47] J. D. Murry, *Asymptotic Analysis* (Springer-Verlag, Berlin, 1984).