Statistical model for receptor-ligand binding thermodynamics

Marc Thilo Figge*
Centre for Theoretical Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
(Received 14 August 2002; published 6 December 2002)

We present a simple statistical model to describe receptor-ligand binding in terms of the number of binding contact residues and the number of separate binding regions as a function of the temperature. The fact that the binding depends on various random factors is modeled by a distribution of local binding energies and we take into account that the interaction between receptor and ligand is only of significance for the activation of the receptor if the total binding energy exceeds a threshold energy. We interpret our results in the light of both experimentally observed antibody-antigen binding configurations and theoretical studies in the zero-temperature limit.

DOI: 10.1103/PhysRevE.66.061901 PACS number(s): 87.15.Aa, 87.15.Kg, 05.50.+q

I. INTRODUCTION

During recent years, there has been a growing interest in the research directed towards the understanding and control of molecular recognition in biochemical processes. Molecular recognition is realized by the specific binding of ligands to receptors. In the course of this process, the receptors become activated and may initiate a cascade of events [1]. A well-known example of a molecular system that is controlled by receptor-ligand binding is the immune system, which performs the important task of defending living organisms against malignantly transformed cells and pathogenic organisms. This is realized by immune responses that are mounted by the specific binding of antibody to antigen or T-cell receptor to major histocompatibility complex (MHC) molecule [2]. From a pharmacologic point of view, the understanding and control of molecular recognition is an important issue in the context of drug design. This includes, for instance, the design of ligands which are supposed to bind highly specific only to their target cell receptors in order to prevent undesired side effects, or which should be particularly long-living and able to bind the receptors of tumor cells in order to suppress their interactions with other cells.

In this paper, we study thermodynamic properties of the receptor-ligand binding within a statistical binding model and discuss our results in the light of experiments on antibody-antigen binding. X-ray crystallographic measurements have revealed that the binding between the receptors of the immune system and their ligands extends over an area of roughly 600 $Å^2$ consisting of several linear binding regions [3–7]. In particular, for an antibody-antigen (lysozyme) complex the receptor-ligand overlap region consists typically of 100 amino acids. The receptor-ligand binding area of such a complex has been found in one experiment to consist of a single linear binding region formed by 17 amino acids [3] and in another experiment to involve 14 amino acids distributed over three separate linear binding regions [4]. Similar sizes for the binding area have not only been confirmed for another type of antibody-antigen complex with a single linear binding region of 15 amino acids [6], but also for a binding complex of T-cell receptor and MHC molecule [7].

The existence of a typical size for the binding region has been successfully interpreted within a probabilistic model as the optimal result of efficient self–nonself discrimination by antibodies and T-cell receptors [8]. In this model, the receptor (antibody) and ligand (antigen) are represented by strings, as is schematically shown in Fig. 1. Antibodies and antigens are composed of amino acids that belong to different complementarity classes. Based on empirical data, amino acids are distinguished as being hydrophobic, or hydrophilic, and if hydrophilic as being positively or negatively charged [8,9]. This gives rise to $c=3$ complementarity classes, where hydrophobic is complementary to hydrophobic, positively charged is complementary to negatively charged, and vice versa. In the probabilistic model, it is assumed that antibodies and antigens are randomly taken from finite repertoires, respectively, of size n_R and n_L. The probability that an antibody repertoire of size n_R has the property that each of the n_L different antigens is recognized by at least one antibody has been calculated under the condition that none of the organism's n_S self-molecules is recognized by any of these antibodies [8]. The result can be expressed in terms of a number r, which denotes the ratio between the size of the receptor-ligand binding region and the size of the receptor-ligand overlap region. In agreement with the experiments [3–7], the ratio is found to be of intermediate order of magnitude: $r \sim 15\%$ [8]. That this value of r reflects optimal antigen recognition by a finite antibody repertoire size can be understood as follows: If each antigen would be recognized

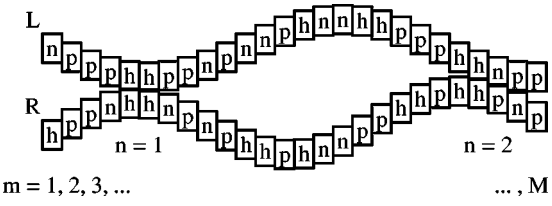

FIG. 1. Schematic representation of antibody (R) and antigen (L) as strings which are composed of amino acids that belong to three different complementarity classes: hydrophobic (h), hydrophilic positively charged (p), and hydrophilic negatively charged (n). See the text for details.

*Email address: m.t.figge@phys.rug.nl

over the entire overlap region, corresponding to $r \sim 100\%$, one specific antibody would be required for each antigen, which is not consistent with typical repertoire sizes. On the other hand, if antibodies would recognize too short amino acid patterns, say of the order $r \sim 1\%$, they could as well bind to self-molecules.

A different route to model receptor-ligand binding is followed in the present paper, where we calculate the (free) energy of the receptor-ligand complex and obtain a particular binding configuration as the result of a minimization procedure. The statistical element in our model originates from the fact that the local binding energy between two contact residues of receptor and ligand is not the same for any such pair. It rather depends on various random factors, such as the local environment, the influence of neighboring contact residues, and conformational effects of the binding. We take this into account by a distribution of local binding energies. A second feature of our model is that the interaction between receptor and ligand is only considered to be of significance for the activation of the receptor if the binding energy exceeds a certain threshold value. We refer to this value as threshold energy. It is associated with the energy required to stabilize a binding region against the steric interaction at its end points, where binding and nonbinding pairs of contact residues are located next to each other, and against the thermally induced motion between receptor and ligand that is counteracting the binding. Recently, a similar model has been used to calculate the polymer-dimer binding probability as a function of the threshold energy for different binding energy distributions [10].

The paper is organized as follows: We introduce the binding model in Sec. II and show in Sec. III that the formation of a binding configuration occurs with a finite probability. Next, in Sec. IV, we map the binding model on a model of the random-field Ising-type and calculate the corresponding free energy. This enables us to study thermodynamic properties of the receptor-ligand binding in Sec. V and to interpret our results in the context of experimentally observed antibody-antigen binding configurations. Finally, in Sec. VI, we summarize and conclude this paper.

II. BINDING MODEL

We introduce a binding model where the receptor and ligand are considered to be bound by contact residues that are distributed over several linear binding regions along the receptor-ligand overlap region. This is indicated in Fig. 1, where we label the contact residues of the receptor by $m = 1, \ldots, M$ with M the total number of receptor contact residues accessible to the ligand. The nth binding region ($n = 1, \ldots, N$) is defined by its two ending contact residues, respectively, $m_{2n-1}+1$ and m_{2n}, where we assume, without loss of generality, the ordering $m_1+1 < m_2 < m_3+1 < m_4 < \cdots < m_{2N-1}+1 < m_{2N}$. We will refer to two binding contact residues as a pair of binding contact residues and their number is simply given by $M_b = \Sigma_{n=1}^{N}(m_{2n} - m_{2n-1})$.

The binding model is defined by the energy associated with a receptor-ligand binding configuration and can be written as the sum of two contributions,

$$E = E_{\parallel} + \Delta E_{\sharp}. \tag{1}$$

The energy of the unbound receptor and ligand is represented by $E_{\parallel}$ and does not depend on M_b and N, while the energy change due to the formation of a particular receptor-ligand binding configuration is given by

$$\Delta E_{\sharp} = E_t + \sum_{n=1}^{N} \sum_{m=m_{2n-1}+1}^{m_{2n}} E_b(m). \tag{2}$$

Here, E_t denotes the threshold energy while the second term represents the binding energy since the sum runs over all the M_b pairs of binding contact residues with local binding energy $E_b(m)$.

The local binding energy depends on various factors, e.g., the type of interaction between two contact residues, the influence of neighboring contact residues, conformational effects of the binding, and also the environment. Because the precise distribution of local binding energies is usually not known, we will assume throughout this paper that $E_b(m)$ is Gaussian distributed. We expect that, as far as the qualitative conclusions are concerned, our results will not depend on this particular choice for the distribution, which has the technical advantage that properties of the receptor-ligand binding can be calculated analytically. The corresponding probability density,

$$f(E_b(m)) = \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(-\frac{[E_b(m) - \overline{E_b}]^2}{2\varepsilon^2} \right), \tag{3}$$

is characterized by the average local binding energy,

$$\overline{E_b} = \langle E_b(m) \rangle_{E_b}, \tag{4}$$

and by the standard deviation

$$\varepsilon = \{ \langle [E_b(m) - \overline{E_b}]^2 \rangle_{E_b} \}^{1/2}, \tag{5}$$

where $\langle \cdots \rangle_{E_b}$ denotes the Gaussian average with probability density Eq. (3). For a given binding-energy distribution there are M_f contact residue pairs that favor the binding, i.e., the corresponding local binding energy $E_b(m) < 0$. The ratio $\nu \equiv M_f/M$ is the fraction of contact residues that favor the binding and can be expressed in terms of $\overline{E_b}$ and ε,

$$\nu = \int_{-\infty}^{0} f(E_b(m)) dE_b(m) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\overline{E_b}}{\sqrt{2}\varepsilon} \right) \right], \tag{6}$$

where $\mathrm{erf}(x) \equiv (2/\sqrt{\pi}) \int_0^x dz e^{-z^2}$ denotes the error function. This number plays an important role in characterizing how specific a particular binding configuration is, however, whether binding between the receptor and ligand is realized depends on whether this is energetically favorable. In other words, as follows from Eq. (1), receptor-ligand binding takes place if

$$\Delta E_{\sharp} \leqslant 0, \tag{7}$$

meaning that the threshold energy is compensated by the binding energy.

The threshold energy consists of two parts,

$$E_t = NE_s + M_b E_v(T), \tag{8}$$

where E_s accounts for the steric interactions that arise at the end points of the N binding regions where binding and nonbinding pairs of contact residues are located next to each other. The second contribution in Eq. (8) accounts for the thermally induced relative motion between binding contact residue pairs. The vibrations counteract the binding and ultimately cause the dissociation of the receptor and ligand at sufficiently high temperatures T. This is modeled by the energy contribution $E_v(T)$, which is a function that increases with the temperature from its zero-temperature value $E_v(T=0)=0$ and will be specified below on phenomenological grounds. The binding and the nonbinding pairs of contact residues enter into the energy expression Eq. (1) with thermal energy contributions of opposite sign, respectively, $+M_b E_v(T)/2$ and $-(M-M_b)E_v(T)/2$, so that the energy of the unbound receptor and ligand becomes

$$E_{\|} = -\frac{M}{2} E_v(T). \tag{9}$$

It is instructive to calculate the probability for the occurrence of receptor-ligand binding within the statistical model Eq. (1). We do this in Sec. III before we study thermodynamic properties of the receptor-ligand binding in Sec. V.

III. BINDING PROBABILITY

We consider the probability that receptor-ligand binding occurs between M_b pairs of contact residues distributed over N binding regions along an overlap region that consists of M contact residues. Binding between the receptor and ligand is energetically favorable if the condition Eq. (7) is satisfied, however this condition is not sufficient to calculate the binding probability of the optimal binding configuration with the lowest energy $\Delta E_\sharp$. We take this into account in the formal expression of the binding probability,

$$P = \frac{1}{\mathcal{N}} \langle \Theta(-\Delta E_\sharp) \mathcal{C}(m_1, m_2; \ldots; m_{2N-1}, m_{2N}) \rangle_{E_b}, \tag{10}$$

where the step function

$$\Theta(x) = \begin{cases} 1 & \text{for } x \geqslant 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{11}$$

ensures that P satisfies the condition Eq. (7), while the function $\mathcal{C}(m_1, m_2; \ldots; m_{2N-1}, m_{2N})$ serves to guarantee that P is the binding probability of the optimal binding configuration. An expression for $\mathcal{C}(m_1, m_2)$ will be given in Sec. III B, where we consider the case of a single binding region ($N=1$). Finally, $\mathcal{N}$ is determined by the normalization condition,

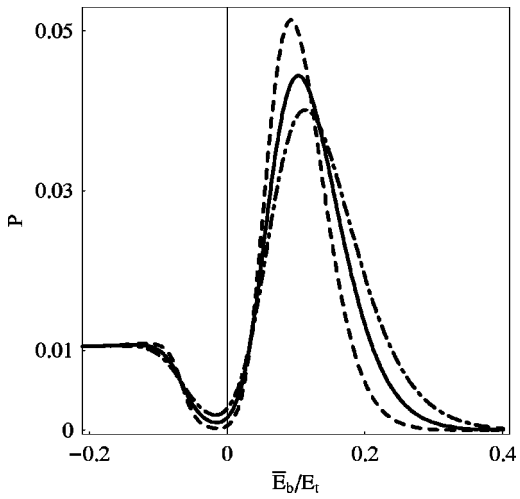

FIG. 2. Probability for the occurrence of receptor-ligand binding by $M_b=15$ contact residue pairs along an overlap region of $M=100$ contact residues as a function of the average local binding energy $\overline{E_b}/E_t$. The three curves correspond to standard deviation $\varepsilon/E_t=0.07$ (dashed line), $\varepsilon/E_t=0.10$ (solid line), and $\varepsilon/E_t=0.13$ (dashed-dotted line) of the binding-energy distribution.

$$\int_0^M P dM_b = 1, \tag{12}$$

with respect to the number M_b of binding contact residue pairs.

A. Occurrence of receptor-ligand binding

We start with the calculation of the probability that, for a given distribution of local binding energies, binding between M_b contact residue pairs occurs at all. In this case $\mathcal{C}(m_1, m_2; \ldots; m_{2N-1}, m_{2N}) = 1$, meaning that all possible binding configurations with binding regions $1 \leqslant N \leqslant M_b$ are taken into account, and Eq. (10) reduces to

$$P = \frac{1}{\mathcal{N}} \left\langle \Theta\left(-E_t - \sum_{n=1}^{N} \sum_{m=m_{2n-1}+1}^{m_{2n}} E_b(m) \right) \right\rangle_{E_b}. \tag{13}$$

We find [11]

$$P = \frac{1}{2\mathcal{N}} [1 - \mathrm{erf}(g(M_b, N))], \tag{14}$$

where we defined

$$g(M_b, N) \equiv \frac{E_t + M_b \overline{E_b}}{\sqrt{2 M_b \varepsilon^2}}. \tag{15}$$

In Fig. 2, we plot the probability Eq. (14) as a function of the average local binding energy $\overline{E_b}/E_t$ for three different values of the standard deviation ε/E_t. We set the number of binding contact residue pairs to $M_b=15$ and consider an overlap region of $M=100$ contact residues. These parameters correspond to realistic values for antibody-antigen binding as observed in crystallographic x-ray experiments [3–6].

At negative average local binding energies, $\overline{E_b}/E_t \ll -1/M_b$, we find that the binding probability becomes independent of the standard deviation ε and approaches the constant value $P=1/M$. This result can also be derived directly from Eq. (14) and has the simple interpretation that any kind of binding configuration can exist in this limit, because the fraction of contact residues that favor the binding between the receptor and ligand tends to $\nu=1$ [see Eq. (6)]. For a vanishing average local binding energy, $\overline{E_b}/E_t=0$, we have $\nu=0.5$ and the probability Eq. (14) is a function of the standard deviation ε of the binding-energy distribution: $P \propto 1-\mathrm{erf}(E_t/\sqrt{M_b \varepsilon^2})$.

Two regimes have to be distinguished for positive average local binding energies. If $\overline{E_b}/E_t \gg \varepsilon/E_t$, the fraction of contact residues that favor the binding between the receptor and ligand approaches $\nu=0$ and, consequently, the probability for the occurrence of binding configurations is found to be strongly suppressed in this case. Most interesting, at positive average local binding energies $\overline{E_b}/E_t < \varepsilon/E_t$, the probability P has a sharp peak, which indicates that binding by $M_b=15$ contact residues is most likely realized if the fraction of contact residues that favor the binding is in the range $0.1<\nu<0.5$. This can be easily understood if one takes into account that antibodies and protein antigens are composed of amino acids that belong to different complementarity classes. Amino acids are distinguished as being hydrophobic or hydrophilic, and if hydrophilic, as being positively or negatively charged [8,9]. This gives rise to $c=3$ complementarity classes, where hydrophobic is complementary to hydrophobic, positively charged, is complementary to negatively charged and vice versa. Neglecting for a moment the effect of all random sources on the binding, we estimate that the fraction of complementary pairs of amino acids along the antibody-antigen overlap region is $\nu=c^{-1}\sim 0.33$. This value is, in fact, within the relevant range for ν. Furthermore, it follows from Eq. (6) that $\overline{E_b}/\varepsilon \sim 0.4$ in this case, so that Eq. (14) may be replaced by

$$P \approx \frac{1}{2\mathcal{N}} \frac{\exp(-g(M_b,N)^2)}{\sqrt{\pi} g(M_b,N)}, \tag{16}$$

which represents a valid asymptotic expression for average local binding energies in the range $\varepsilon/E_t < \overline{E_b}/\varepsilon < 1$. The interpretation of the expression Eq. (16) is that the optimal fluctuation which can induce binding, i.e., the binding-energy fluctuation with the largest weight, has a constant negative value between the M_b pairs of binding contact residues,

$$E_b^*(M_b,N)=\overline{E_b}-\delta E_b^*(M_b,N), \tag{17}$$

and is zero between all other contact residues. The corresponding amplitude is determined from the energy balance, $\Delta E_\sharp=0$, to be

$$\delta E_b^*(M_b,N)=\overline{E_b}+E_t/M_b \tag{18}$$

and the weight of the optimal binding-energy fluctuation,

$$f(E_b^*(M_b))^{M_b} \propto \exp\left(-\frac{M_b \delta E_b^*(M_b)^2}{2\varepsilon^2}\right), \tag{19}$$

is precisely the exponential factor appearing in Eq. (16). Thus, all binding-energy fluctuations that contribute significantly to the binding probability are close to the optimal fluctuation. For the typical number of binding contact residue pairs,

$$M_b^*=\frac{NE_s}{E_v(T)+\overline{E_b}}, \tag{20}$$

the weight Eq. (19) reaches its maximal value,

$$f(E_b^*(M_b^*))^{M_b^*} \propto \exp\left(-\frac{2\overline{E_b}E_t}{\varepsilon^2}\right), \tag{21}$$

and is exponentially decreasing with increasing threshold energy E_t. Furthermore, Eq. (20) reveals that the typical number of binding contact residue pairs is the result of a competition, because M_b^* is directly proportional to the steric interaction energy and inversely proportional to the energy associated with the thermally induced vibrations within pairs of binding contact residues.

B. Optimized receptor-ligand binding

We consider a binding configuration that consists of a single binding region ($N=1$), which we denote by $\{m_1,m_2\}$, involving $M_b=m_2-m_1$ contact residue pairs. The condition Eq. (7) reduces to

$$\Delta E_\sharp[\{m_1,m_2\}]=E_t+\sum_{m=m_1+1}^{m_2} E_b(m) \leqslant 0, \tag{22}$$

with

$$E_t=E_s+M_b E_v(T). \tag{23}$$

To determine the probability P of the optimal binding configuration $\{m_1,m_2\}$ with respect to all other configurations $\{m_1',m_2'\}$ that also consist of a single binding region, we have to calculate Eq. (10) with

$$\mathcal{C}(m_1,m_2)=\prod_{m_1'<m_2'} \Theta(\Delta E_\sharp[\{m_1',m_2'\}] - \Delta E_\sharp[\{m_1,m_2\}]). \tag{24}$$

Since $\mathcal{C}(m_1,m_2)$ does not account for binding configurations with more than one binding region, the corresponding binding probability will only be valid in a range of the parameters $\overline{E_b}$ and ε, for which binding configurations with $N=1$ will be the optimal ones. This restriction simplifies the calculation of the binding probability P considerably. We present the calculation in the Appendix, where we arrive at the final expression,

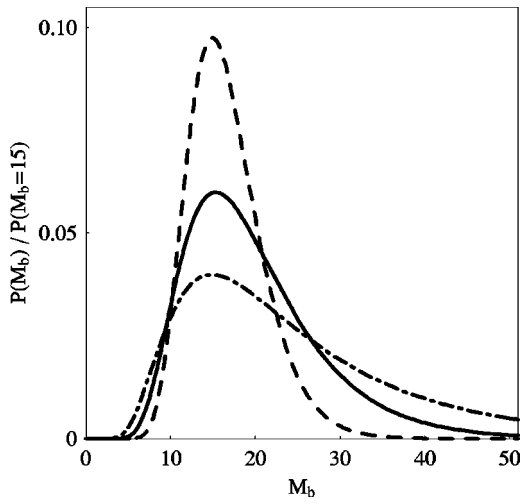

FIG. 3. Probability for receptor-ligand binding by a single binding region ($N=1$) as a function of the number M_b of binding contact residue pairs ($M=100$). The three curves correspond to three parameter sets $\{\overline{E_b}/E_t, \varepsilon/E_t\}$ for which P is maximal at $M_b=15$: {0.062,0.07} (dashed line), {0.05,0.10} (solid line), and {0.035,0.13} (dashed-dotted line).

$$P=\frac{\mathcal{K}(M)\overline{E_b}^2\sqrt{M_b}}{\varepsilon(E_t+M_b\overline{E_b})}\left[\frac{Z\left(\frac{E_t}{M_b\varepsilon}\right)}{Z\left(\frac{\overline{E_b}}{\varepsilon}\right)}\right]^2 \times\exp\left(-\frac{(E_t-M_b\overline{E_b})^2}{2M_b\varepsilon^2}\right), \tag{25}$$

with $Z(x)=\tanh(1.14x)$ and

$$\mathcal{K}(M)=\sqrt{\frac{8}{\pi}}\left\{\operatorname{erf}\left[\frac{g(M_b^*,1)}{2}\left(\frac{M}{M_b^*}-1\right)\right]+\operatorname{erf}\left(\frac{g(M_b^*,1)}{2}\right)\right\}^{-1} \tag{26}$$

contains M_b^* as defined by Eq. (20).

The binding probability P is dominated by the exponential factor which is just the normalized Gaussian weight of the optimal binding energy fluctuation,

$$P\propto\frac{f(E_b^*(M_b))^{M_b}}{f(E_b^*(M_b^*))^{M_b^*}}, \tag{27}$$

as follows from Eqs. (19) and (21) with $E_b^*(M_b)=-E_t/M_b$ [see Eqs. (17) and (18)]. In Fig. 3, we plot the binding probability Eq. (25) for three different parameter sets $\{\overline{E_b}/E_t, \varepsilon/E_t\}$ as a function of the number M_b of binding contact residue pairs. For each set of parameters, the corresponding binding probability P has its maximal value at $M_b=15$ and the binding probability falls off exponentially for smaller ($M_b<15$) or larger ($M_b>15$) binding regions. Similarly, as can be deduced from Eq. (25), the binding probability has a peak structure as a function of the threshold energy E_t. It is intuitively clear that the binding probability P decreases with increasing threshold energy E_t, as has also been found from calculations of the polymer-dimer binding probability [10]. However, a less obvious result is that the binding probability decreases exponentially for decreasing threshold energies $E_t<M_b\overline{E_b}$. This behavior indicates once again that all binding-energy fluctuations that contribute significantly to the receptor-ligand binding probability are close to the optimal binding-energy fluctuation.

We note that, while a rigorous calculation of the general binding probability $P=P(M,M_b,N)$ is quite involved, it may be estimated on the basis of the expression Eq. (25) for $P(M,M_b,1)$. This is done by dividing the receptor-ligand overlap region into N segments each containing a single binding region. Treating the N segments as independent units, we may write

$$P(M,M_b,N)\approx\prod_{n=1}^{N}P(M(n),M_b(n),1), \tag{28}$$

where $M(n)$ and $M_b(n)$ are the number of contact residues that are, respectively, accessible and binding in the nth segment. We assume here that $M(n)\gg M_b(n)$ holds in each segment. Since different binding-energy distributions give rise to clearly distinct receptor-ligand binding configurations associated with the corresponding optimal binding energy fluctuations, we may conclude that the parameter set $\{\overline{E_b}/E_t, \varepsilon/E_t\}$ which maximizes the binding probability $P(M,M_b,N)$ prefers an equal distribution of binding contact residues with $M_b^*(n)\sim M_b^*/N$. Within this simplified picture, we obtain a condition for the occurrence of receptor-ligand binding by a rough estimate as follows: Assuming that the typical number $M_b^*(n)$ is mainly determined by the exponential factor in Eq. (25), we find from Eq. (20) that $M_b^*(n)\sim E_s/[E_v(T)+\overline{E_b}]$. It thus follows that the probability for the occurrence of receptor-ligand binding becomes zero for $E_v(T)\sim E_s-\overline{E_b}$, which can be related to the temperature at which the receptor and ligand dissociate.

To summarize, the calculation of the binding probability within the statistical model Eq. (1) yields the qualitative result that distinct binding configurations occur at sufficiently low temperatures with a finite probability.

IV. BINDING FREE ENERGY

In order to study thermodynamic properties of the receptor-ligand binding, we calculate the free energy of the binding model. This can be done by rewriting Eq. (1) in terms of a local binding variable, σ_m, which describes the state of contact residue m by one of its two possible values,

$$\sigma_m=\begin{cases}-1 & \text{if binding,}\\ +1 & \text{if nonbinding.}\end{cases} \tag{29}$$

The energy E of a particular binding configuration $\{\sigma_m\}$ can again be written as the sum of two parts,

$$E=E_0+E_1[\{\sigma_m\}]. \tag{30}$$

Here, the first part,

$$E_0=\frac{1}{2}\sum_{m=1}^{M}E_b(m), \tag{31}$$

is independent of the binding variable σ_m, while the second part is given by

$$E_1[\{\sigma_m\}]=\frac{1}{2}\sum_{m=1}^{M}\left[E_s\frac{(1-\sigma_m\sigma_{m+1})}{2}-E_v(T)\sigma_m-E_b(m)\sigma_m\right]. \tag{32}$$

The first two terms represent the threshold energy Eq. (8), where the steric interactions are represented by the first term since the sum over

$$\frac{1}{2}(1-\sigma_m\sigma_{m+1})=\begin{cases}0 & \text{if } \sigma_m=\sigma_{m+1}\\ 1 & \text{if } \sigma_m=-\sigma_{m+1}\end{cases} \tag{33}$$

counts the number of binding-region end points within the binding configuration $\{\sigma_m\}$. The second term accounts for the effect of the thermal vibrations which favor the nonbinding state, $\sigma_m=+1$, because $E_v(T>0)>0$. Finally, the third term in Eq. (32) corresponds to the local binding energy, $E_b(m)\equiv\overline{E_b}+\varepsilon s_m$, where s_m is a random variable that is Gaussian distributed according to the probability density

$$f(s_m)=\frac{\exp\left(-\frac{1}{2}s_m^2\right)}{\sqrt{2\pi}}. \tag{34}$$

We note that Eq. (32) may be interpreted as a one-dimensional random-field Ising model, where σ_m denotes an Ising spin at site m of a chain and can be either up ($\sigma_m=+1$) or down ($\sigma_m=-1$) due to the presence of a (quite exotic) magnetic field that has a random component and a temperature-dependent component.

The average free energy may be written as

$$\langle F\rangle_s=\langle F_0\rangle_s+\langle F_1\rangle_s, \tag{35}$$

where $\langle\cdots\rangle_s$ denotes the Gaussian average with probability density Eq. (34). The first term originates from the energy contribution E_0,

$$\langle F_0\rangle_s=\frac{M}{2}\overline{E_b}. \tag{36}$$

The second term is given by the logarithm of the Boltzmann-weighted sum over all possible binding configurations with energy $E_1[\{\sigma_m\}]$,

$$\langle F_1\rangle_s=-T\left\langle\ln\sum_{\{\sigma_m\}}e^{-E_1[\{\sigma_m\}]/T}\right\rangle_s, \tag{37}$$

in units where Boltzmann's constant $k_B=1$. It is straightforward to compute $\langle F_1\rangle_s$ applying the transfer-matrix formalism, where the sum in Eq. (37) is written as an ordered product of matrices $\mathcal{T}_m$:

$$\sum_{\{\sigma_m\}}e^{-E_1[\{\sigma_m\}]/T}=\sum_{\sigma,\sigma'=\pm1}\langle\sigma'|\prod_{m=1}^{M}\mathcal{T}_m|\sigma\rangle. \tag{38}$$

Here, the final summation over σ,σ' accounts for all possible boundary conditions for the binding variable at the ends of the receptor-ligand overlap region. On the basis spanned by the two vectors $|+\rangle$ and $|-\rangle$ (corresponding, respectively, to $\sigma=+1$ and $\sigma=-1$), the transfer matrix $\mathcal{T}_m$ reads

$$\mathcal{T}_m=\begin{pmatrix}e^{+[E_v(T)+E_b(m)]/2T} & e^{-[E_s-E_v(T)-E_b(m)]/2T}\\ e^{-[E_s+E_v(T)+E_b(m)]/2T} & e^{-[E_v(T)+E_b(m)]/2T}\end{pmatrix} \tag{39}$$

and is a function of the mth contact residue through the local binding energy $E_b(m)=\overline{E_b}+\varepsilon s_m$. It is clear from Eq. (38) that the numerical calculation of the free energy $\langle F_1\rangle_s$ simply requires the repeated multiplication of a matrix with a vector.

Recently, we derived a closed analytical expression for the free energy $\langle F_1\rangle_s$ which is valid in the low-temperature limit. The tedious calculation is presented in Ref. [12] and will not be reproduced here. It requires the derivation of the continuum version of the discrete model Eq. (32), which is found to describe the relaxation of a single quantum spin-$\frac{1}{2}$ in a magnetic field. The corresponding Hamiltonian can be mapped onto a Brownian motion model which is defined in terms of a Langevin equation. Solving the corresponding Fokker-Planck equation yields a probability distribution which is used to perform the average of the free energy Eq. (37) analytically. For temperatures $T\ll E_s$ and energies $\overline{E_b},\varepsilon,E_v(T)\ll E_s$, we obtain to leading order in the number M of contact residues the expression [12]

$$\langle F\rangle_s=\frac{M}{2}\left[\overline{E_b}-\frac{E_v(T)+\overline{E_b}}{\tanh(\varrho)}\right], \tag{40}$$

where the dimensionless variable

$$\varrho=\frac{(E_v(T)+\overline{E_b})}{\varepsilon^2}\tilde{E}_s(T,\varepsilon) \tag{41}$$

contains the effective steric interaction energy

$$\tilde{E}_s(T,\varepsilon)=\begin{cases}E_s+4k\varepsilon & \text{for } T\leqslant T_0\\ E_s-4T\ln\left(\frac{T}{ek\varepsilon}\right) & \text{for } T>T_0.\end{cases} \tag{42}$$

Here, $T_0=k\varepsilon$ and $k=e^{-(\gamma+2)/2}/\sqrt{2}\simeq0.195$ contains Euler's constant $\gamma\simeq0.577$. The temperature T_0 enters the expression for the free energy in the course of deriving the model's continuum version [12]. The reason is that in the continuum model, the end points of a binding region can take any posi-

tion along the receptor-ligand overlap region and are not restricted to the discrete locations of the contact residues in the original discrete model. In other words, each end point of a binding region enters in the expression Eq. (37) with a weight

$$e^{-\tilde{E}_s(T,\varepsilon)/2T} = l(T,\varepsilon)e^{-E_s/2T}. \tag{43}$$

Here, $l(T,\varepsilon)$ is the typical size per length of a contact residue for the thermal fluctuation of the end-point position along the receptor-ligand overlap region. We estimate $l(T,\varepsilon)$ from the criterion that shifting the end position over $l(T,\varepsilon)$ from its optimal position requires a fluctuation of the order of T:

$$\left\langle \left(\sum_m \varepsilon s_m \sigma_m \right)^2 \right\rangle_s^{1/2} = \varepsilon l(T,\varepsilon)^{1/2} \sim T. \tag{44}$$

Therefore, the continuum model gives incorrect results at temperatures below $T_0 \sim \varepsilon$, where the thermal fluctuation of the end-point position and the length of a contact residue are of the same order: $l(T_0,\varepsilon)\sim 1$. This explanation is confirmed by numerical simulations of the discrete model Eq. (32), where for $T<T_0$ the corresponding free energy is in fact only weakly depending on the temperature [12]. In the analytical expression of the free energy this is taken into account by the effective steric interaction energy Eq. (42).

V. BINDING THERMODYNAMICS

The calculation of the free energy enables us to study for a given distribution of local binding energies the corresponding binding configuration as a function of the temperature. It follows directly from Eqs. (35)–(37) that the number of binding regions is obtained by differentiation of the free energy with respect to E_s,

$$N = \frac{\partial}{\partial E_s}\langle F\rangle_s, \tag{45}$$

and the number of binding contact residue pairs by differentiation with respect to $\overline{E_b}$,

$$M_b = \frac{\partial}{\partial \overline{E_b}}\langle F\rangle_s. \tag{46}$$

Here and from now on we denote by N and M_b the corresponding thermally and Gaussian averaged quantities.

Using the analytical expression Eq. (40) for the free energy, we obtain expressions for the number of binding regions,

$$N = \frac{M[E_v(T)+\overline{E_b}]^2}{2\varepsilon^2 \sinh(\varrho)^2}, \tag{47}$$

and for the number of binding contact residue pairs,

$$M_b = \frac{M}{2}\left[1 - \frac{\sinh(2\varrho)-2\varrho}{2\sinh(\varrho)^2}\right]. \tag{48}$$

The solution of Eq. (48) in terms of $\varrho = \varrho(M_b/M)^*$ for a given receptor-ligand binding configuration is easily obtained numerically and can be represented by the series

$$\varrho(M_b/M)^* = c_0 + \sum_{l=1}^{L} c_l \left(\frac{M_b}{M}\right)^{1/l}. \tag{49}$$

For $L=4$ and coefficients $c_0=8.546$, $c_1=-0.107$, $c_2=-21.745$, $c_3=70.135$, and $c_4=-58.005$, the relative deviation of ϱ^* from the exact solution is already well below 1% over the entire range $0\leqslant M_b/M\leqslant 0.5$. Combining Eqs. (41), (47), and (49), we can calculate the average local binding energy and the standard deviation of the binding energy distribution Eq. (3) at temperature $T=0$. We find

$$\overline{E_b} = K(M,M_b,N)E_s \tag{50}$$

and

$$\varepsilon = \frac{\sqrt{M}K(M,M_b,N)}{\sqrt{2N}\sinh(\varrho^*)}E_s, \tag{51}$$

where we defined

$$K(M,M_b,N) \equiv \frac{2N\sinh(\varrho^*)^2}{\varrho^* M - 4\sqrt{2}k\sqrt{MN}\sinh(\varrho^*)}, \tag{52}$$

which is a positive number in the parameter region $N\leqslant M_b \leqslant M/2$.

In order to analyze the specificity of the receptor-ligand binding, we calculate the binding free energy $\langle F(T=0)\rangle_s$ and the fraction of contact residue pairs that favor the binding, $\nu = M_f/M$, for any binding configuration characterized by the parameter set $\{M,M_b,N\}$. It follows from Eqs. (40), (49), and (50) that

$$\langle F(T=0)\rangle_s = \frac{M\overline{E_b}}{2}\left[1 - \frac{1}{\tanh(\varrho^*)}\right], \tag{53}$$

and from Eqs. (6), (50), and (51) that

$$\nu = \frac{1}{2}\left[1 - \mathrm{erf}\left(\sqrt{\frac{N}{M}}\sinh(\varrho^*)\right)\right]. \tag{54}$$

We plot ν and $\langle F(T=0)\rangle_s$ as a function of M_b/M for binding configurations with $0.01\leqslant N/M\leqslant 0.04$, respectively, in Fig. 4 and Fig. 5. It can be seen in Fig. 4 that ν is larger for a larger number of binding contact residue pairs, M_b, and is smaller for a larger number of separate binding regions, N. In Fig. 5, we see that, for a fixed number of binding regions N, a binding configuration is energetically more favorable if more binding contact residue pairs M_b are involved, while keeping the number M_b fixed, the energetically most favorable binding configuration is that with the largest number of separate binding regions $N=M_b$. A criterion to quantify the specificity of a particular receptor and ligand may now be

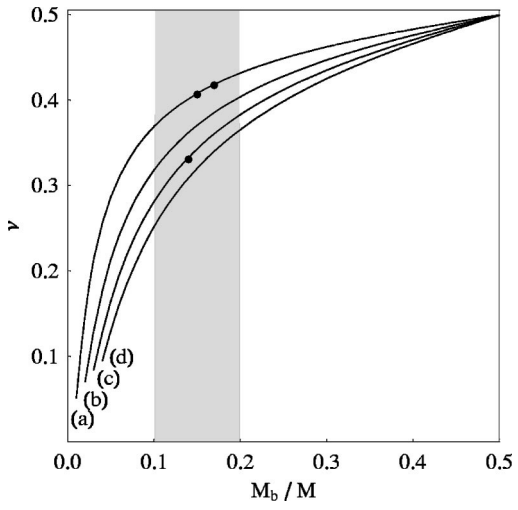

FIG. 4. The number of contact residues $\nu = M_f/M$ that favor the binding as a function of M_b/M for binding configurations with (a) $N/M = 0.01$, (b) $N/M = 0.02$, (c) $N/M = 0.03$, and (d) $N/M = 0.04$. The three dots refer to the three binding configurations $\{M, M_b, N\} = \{100,15,1\}$, $\{100,17,1\}$, and $\{100,14,3\}$. The typical regime for antibody-antigen binding configurations is indicated by the shaded region.

formulated as follows: The larger the typical energy contribution to the binding free energy per contact residue pair that favors the binding is, the higher is the specificity of the receptor and ligand. This energy contribution,

$$E_f \equiv \frac{\langle F(T=0)\rangle_s}{M_f}, \tag{55}$$

can be calculated from Eqs. (53) and (54).

We discussed in Sec. III A that for antibody-antigen binding in the absence of any random source, ν is simply given by the inverse of the number of complementarity classes, $c = 3$, and thus $\nu \sim 0.33$. It can be seen in Fig. 4 that the range $0.25 \leqslant \nu \leqslant 0.45$ corresponds to values of the ratio $r \equiv M_b/M$ that are of intermediate order of magnitude: $0.1 \leqslant r \leqslant 0.2$ (shaded region in Fig. 4). This parameter range has been interpreted in Ref. [8] within a probabilistic model as the optimal result of efficient self–nonself discrimination by the receptors of the immune system. Our model not only confirms these conclusions, but also relates the parameter r to the underlying distribution of binding energies. It follows from Eqs. (50)–(54) that for $r \sim 0.5$ the characteristic features for the binding-energy distribution are a relatively small standard deviation ε and an average local binding energy $\overline{E_b} \sim 0$. In other words, in this regime it is equally likely for receptor-ligand contact residue pairs to bind or not to bind. For even larger values of r, the average local binding energy is negative, $\overline{E_b} < 0$, and if antibody-antigen binding would typically occur in this regime, it would be highly specific. However, the finite size of receptor repertoires cannot be reconciled with the limit $r \sim 1$, where one particular antibody is required for each antigen. This leads to the conclusion that receptor-ligand binding in the immune system is more efficiently realized by the binding of antibodies to several, slightly different kinds of antigens involving fewer binding contact residue pairs ($r \ll 1$). On the other hand, for $r \ll 0.1$ the underlying binding-energy distribution is characterized by a standard deviation which is much larger than the average local binding energy, $\varepsilon \gg \overline{E_b}$. This means that randomness plays a dominant role in this limit, so that binding is realized by only a few contact residue pairs that are able to stabilize a binding configuration due to relatively large local binding energies $E_b(m) \sim -E_s$. Clearly, if $r \ll 0.1$ were the typical parameter regime for the receptors of the immune system to function, they would bind equally well to the organism's self-molecules. That this does not happen under healthy conditions may be explained by arguing that the random factors do not govern the binding between receptors and ligands of the immune system, meaning that large local binding energies with $E_b(m) \sim -E_s$ are unlikely to occur. In Table I, we present the values of binding configuration parameters as obtained by analytical and numerical calculations for three experimentally observed antibody-antigen binding configurations [3,4,6]. The parameters of both calculations show the same qualitative behavior and are even in quantitative agreement in the limit $\overline{E_b}, \varepsilon \ll E_s$, where we expect our analytical calculation to be valid. We conclude that the receptor and ligand of the binding configuration $\{M, M_b, N\} = \{100,14,3\}$ represent the most specific combination, since the energy contribution E_f to the binding free energy is the largest of the three binding configurations. Comparing the binding configurations $\{100,15,1\}$ and $\{100,17,1\}$, the specificity of the receptor and ligand is found to be only slightly higher for the latter.

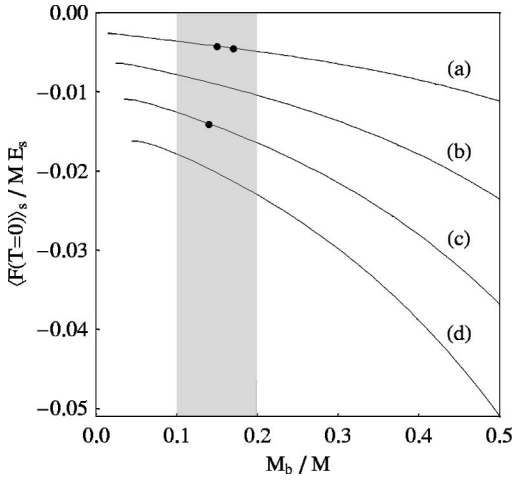

FIG. 5. The binding free energy $\langle F(T=0)\rangle_s$ as a function of M_b/M for binding configurations with (a) $N/M = 0.01$, (b) $N/M = 0.02$, (c) $N/M = 0.03$, and (d) $N/M = 0.04$. The three dots refer to the three binding configurations $\{M, M_b, N\} = \{100,15,1\}$, $\{100,17,1\}$, and $\{100,14,3\}$.

We now turn to the discussion of our results at finite temperatures. The temperature dependence of the energy $E_v(T)$ associated with thermal vibrations of the receptor and ligand is obtained by arguing in a phenomenological way as follows: At temperatures well below the steric interaction energy, $0 \leqslant T \leqslant T_1 \ll E_s$, pairs of binding contact residues may

TABLE I. Parameters for three different binding configurations as obtained from analytical and numerical calculations at $T=0$ for experimentally observed antibody-antigen binding.

Binding configuration			Analytical calculation					Numerical calculation				
M	M_b	N	$\overline{E_b}/E_s$	ε/E_s	ν	$\langle F\rangle_s/E_s$	E_f/E_s	$\overline{E_b}/E_s$	ε/E_s	ν	$\langle F\rangle_s/E_s$	E_f/E_s
100	15	1	0.050	0.213	0.407	−0.419	−0.010	0.050	0.229	0.414	−0.431	−0.010
100	17	1	0.042	0.204	0.418	−0.447	−0.011	0.042	0.217	0.423	−0.457	−0.011
100	14	3	0.187	0.434	0.333	−1.405	−0.042	0.213	0.507	0.337	−1.607	−0.048

slowly move together but the relative motion of contact residues against each other can be neglected, so that $E_v(T \leqslant T_1)$ must be vanishingly small. The value of the temperature T_1 may depend on various factors, e.g., on the mass of the contact residue pairs, and indicates the temperature at which the relative motion between binding contact residues becomes relevant. For temperatures $T>T_1$, the vibrations may initially be described by small harmonic oscillations in the distance between contact residues of the same pair. As these oscillations take place in the plane perpendicular to the direction of the receptor-ligand binding region, we estimate for the corresponding temperature range that $E_v(T)\sim 4T$, so that for the unbound receptor and ligand $E=E_{\|}\sim 2MT$ in accordance with the equipartition theorem. The thermally induced vibrations are only harmonic up to a certain temperature, above which anharmonic effects and significant structural changes of the receptor-ligand complex will become important. However, as a starting point we will assume that harmonic oscillations represent a good approximation. A function that captures the essential behavior described above is given by

$$E_v(T)=\Theta(T-T_1)\frac{4E_s}{\alpha}\ln\left(\cosh\left[\frac{\alpha(T_1-T)}{E_s}\right]\right), \qquad (56)$$

where $\Theta(T-T_1)$ denotes the step function [see Eq. (11)] and the dimensionless parameter α determines how fast $E_v(T)$ evolves into the linear-temperature regime. Taking 10^4 different binding-energy realizations into account, we performed numerical calculations for the parameters $T_1=0.1E_s$ and $\alpha=5$, so that $E_v(T)\propto T$ for $T>0.3E_s$. We checked that qualitatively our results do not depend on this particular choice.

We plot the results of our numerical calculations for the number of binding contact residue pairs and the number of separate binding regions in Figs. 6, 7, and 8, respectively, for the zero-temperature binding configurations $\{M,M_b,N\}=\{100,15,1\}$, $\{100,17,1\}$, and $\{100,14,3\}$. Several common features can be observed in the temperature dependence of the three binding configurations. When the temperature is increased from $T=0$, an increase of M_b and N is observed in each binding configuration. This is a consequence of the thermal fluctuation of the binding regions' end-point positions along the receptor-ligand overlap region. It follows from Eq. (44) that the typical size for the thermal fluctuation scales with $l(T,\varepsilon)\sim T^2/\varepsilon$ and effectively corresponds to a decrease of the steric interaction energy according to Eq. (42). As a consequence, the binding configuration rearranges by increasing M_b and N in order to minimize the binding free energy. This is only possible as long as the counteracting thermal vibrations within contact residue pairs

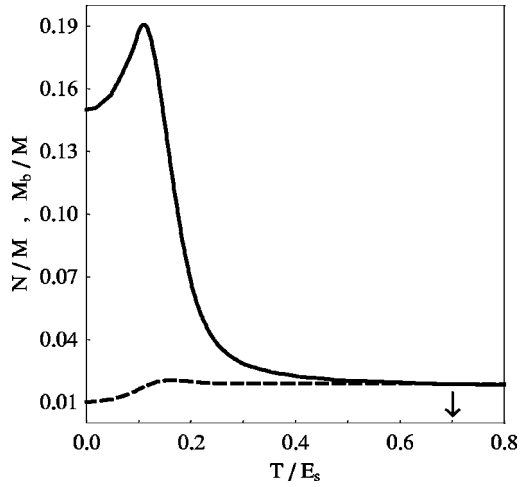

FIG. 6. The number of binding contact residue pairs M_b/M (solid line) and the number of separate binding regions N/M (dashed line) as a function of the temperature T for the zero-temperature binding configuration $\{M,M_b,N\}=\{100,15,1\}$. The arrow indicates that $M_b=N$ at temperature $T\sim 0.70E_s$.

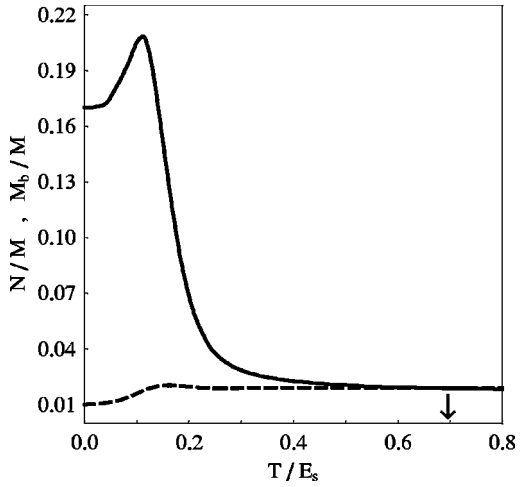

FIG. 7. The number of binding contact residue pairs M_b/M (solid line) and the number of separate binding regions N/M (dashed line) as a function of the temperature T for the zero-temperature binding configuration $\{M,M_b,N\}=\{100,17,1\}$. The arrow indicates that $M_b=N$ at temperature $T\sim 0.69E_s$.

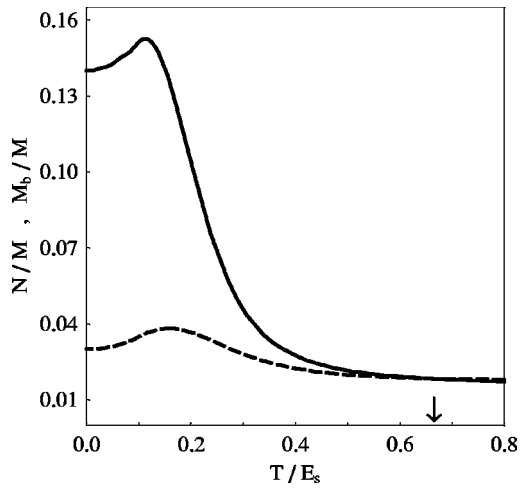

FIG. 8. The number of binding contact residue pairs M_b/M (solid line) and the number of separate binding regions N/M (dashed line) as a function of the temperature T for the zero-temperature binding configuration $\{M,M_b,N\}=\{100,14,3\}$. The arrow indicates that $M_b=N$ at temperature $T\sim 0.66E_s$.

are small. The inset of thermal motion between contact residues of binding pairs is seen at temperature $T\sim T_1$, where the increase of M_b and N becomes weaker. Then, M_b and N reach a maximum before they start to decrease until finally, at a temperature which is indicated by the arrow, the binding configuration is characterized by $M_b=N$. Increasing the temperature further would result in the dissociation of the receptor and ligand since binding configurations with $M_b/N<1$ cannot exist.

It has been suggested that the affinity of the receptor-ligand interaction may be related to the number of binding contact residue pairs [9]. We note that, if a minimum number $M_b\gg 1$ is required to trigger the receptor, its proper functioning may stop already well below the dissociation temperature. Since N is seen to depend only weakly on the temperature, a binding configuration which is most suited to function even at high temperatures is characterized by a large number N at zero temperature, because then $M_b=N$ close to the dissociation temperature may be still sufficiently large to keep the receptor activated. It can be seen from Figs. 6–8 at intermediate temperatures that M_b decreases by roughly a factor 2 slower for the binding configuration with $N=3$ as compared to the binding configurations with $N=1$. Inspection of Eq. (48) reveals that, in general, the decrease of M_b is smaller if the zero-temperature binding configuration is characterized by a larger N and a smaller M_b. Thus, although a zero-temperature binding configuration with a larger number M_b is considered to be more specific, this is not advantageous at finite temperatures since a binding configuration with smaller number M_b can resist the destroying effect of thermal motion better. This can be easily understood from a comparison of two binding configurations: The binding configuration with larger number M_b has a lower binding (free) energy at $T=0$, however, due to thermally induced vibrations its free energy increases faster with the temperature, and the binding configuration can only survive if the number M_b is sufficiently reduced. Similarly, it can be understood that a binding configuration with a larger number N can retain a larger number M_b up to higher temperatures because it has the freedom to rearrange and to gain energy by decreasing the number N. To summarize, at finite temperatures our model predicts that a high specificity of the receptor and ligand is only preferable in terms of a large ratio N/M_b but not in terms of a large ratio M_b/M. This may explain why the immune system does not aim to realize receptor-ligand binding by configurations with $r=M_b/M\sim 1$.

We finally turn to a quantitative discussion of our results. A typical value for the binding free energy requires an estimate of the energy E_s, which sets the energy scale of the model. Because of the fact that the binding configuration of an antibody-antigen complex starts to change significantly before it becomes unstable at temperatures above $T\sim 310$ K, we may estimate from Figs. 6–8 that $T=300$ K corresponds to $T\sim 0.2E_s$, so that $E_s\sim 1500$ K~ 3 kcal/mol. In these units, the binding free energy at $T=300$ K for the zero-temperature binding configurations $\{M,M_b,N\}=\{100,15,1\}$, $\{100,17,1\}$, and $\{100,14,3\}$ is computed to be $\langle F(T=300\ \mathrm{K})\rangle_s=-16.14$ kcal/mol, $\langle F(T=300\ \mathrm{K})\rangle_s=-16.11$ kcal/mol, and $\langle F(T=300\ \mathrm{K})\rangle_s=-19.23$ kcal/mol, respectively. These values represent the correct order of magnitude which is typically measured in biomolecular reactions (-5 to -20 kcal/mol). Furthermore, using the values for $\overline{E_b}$ and ε as given in Table I, also the typical value of the binding energy for contact residues that favor the binding,

$$\overline{E}\equiv \nu^{-1}\int_{-\infty}^{0} E_b(m)f(E_b(m))dE_b(m), \tag{57}$$

is found for all three binding configurations to be of the correct order of magnitude: $\overline{E}\sim -1$ kcal/mol. It should be noted that our model predicts a typical value for the dissociation temperature of the receptor and ligand, $T\sim 1000$ K, which is too large by about a factor 3. This indicates, as expected, that the dissociation of the receptor and ligand cannot simply be described by small thermal vibrations according to Eq. (56). It is obvious that the predicted dissociation temperature would be strongly decreased if significant changes in the structure of the receptor-ligand complex would be taken into account by a suitable modification of $E_v(T)$. However, we note that changing $E_v(T)$ within our model would affect different binding configurations in a similar way and will not influence our qualitative conclusions.

VI. SUMMARY AND CONCLUSIONS

In this paper, we studied thermodynamic properties of receptor-ligand binding within a simple statistical model. The main ingredients of our model are a distribution of local binding energies to account for the effect of various random sources on the binding, and an energy threshold for the total binding energy associated with steric interactions and thermally induced vibrations within the receptor-ligand complex. The calculation of the corresponding binding probability revealed that distinct binding configurations do occur depending on the parameters of the binding-energy distribution. Af-

ter mapping the model on a random-field Ising model, the binding free energy could be calculated, which enabled us to study the thermodynamics of binding configurations in terms of the number of binding contact residue pairs, M_b, and the number of separate binding regions, N.

In contrast to the configurational complexity of realistic receptor-ligand binding configurations, our model is extremely simple, probably the main simplification being the projection of the three-dimensional, entangled structures onto a one-dimensional, effective model. In particular, the influence of thermally induced changes in the receptor-ligand complex is an important issue that has to be addressed in future research. Nevertheless, considering our results in the light of antibody-antigen binding, we not only find that they are in convincing agreement with those of previous zero-temperature studies, but that they allow us to formulate a plausible interpretation of receptor-ligand binding as it is realized in the immune system. We find that the ratio between the size of the receptor-ligand binding region and the size of the receptor-ligand overlap region, $r=M_b/M$, is a consequence of two competing factors. On the one hand, for small $r\sim 1\%$ large local binding energies are required to exceed the threshold energy and to stabilize a binding configuration. In particular, for receptor-ligand binding in the immune system, local binding energies of that order may be unlikely to occur, so that binding over short binding regions is suppressed and self–nonself discrimination is realized in this way. On the other hand, for large $r \gg 10\%$ the binding would be highly specific and a large receptor repertoire would be required. However, our model predicts that binding configurations with a smaller number $r=M_b/M \ll 100\%$ can resist the destroying effect of thermal motion better, so that a high specificity of the receptor and ligand at finite temperatures is only preferable in terms of a large ratio N/M_b but not in terms of a large ratio M_b/M. We conclude that the immune system realizes receptor-ligand binding in an efficient way by a finite number of different receptors that are able to bind several, slightly different kinds of antigens. The binding configurations with $r=M_b/M$ of intermediate order of magnitude reflect a high specificity of receptor and ligand by a large ratio N/M_b.

We finally note that, although we considered receptor-ligand binding in the immune system as an example throughout this paper, our model is more general and may as well be of relevance to other molecular systems that are controlled by receptor-ligand binding.

ACKNOWLEDGMENT

The author gratefully acknowledges lively and stimulating discussions with N. J. Figge.

APPENDIX: OPTIMIZED RECEPTOR-LIGAND BINDING

The binding probability Eq. (10) is calculated with $\Delta E_\sharp$ and $\mathcal{C}(m_1,m_2)$ as given by Eqs. (22) and (24), respectively.

We start by noticing that P factorizes into two independent parts, $P_{\rm out}$ and $P_{\rm in}$, which account for binding regions of sizes, respectively, larger and smaller than M_b contact residues:

$$P=\mathcal{N}^{-1}P_{\rm out}P_{\rm in}. \tag{A1}$$

It is now convenient to define the dimensionless variables

$$t\equiv E_t/\varepsilon, \tag{A2}$$

$$b\equiv \overline{E}_b/\varepsilon, \tag{A3}$$

$$s_m\equiv [E_b(m)-\overline{E}_b]/\varepsilon, \tag{A4}$$

so that

$$\begin{aligned} P_{out}=&\langle \Theta(b+s_{m_1})\Theta(2b+s_{m_1}+s_{m_1-1}) \\ &\times \Theta(3b+s_{m_1}+s_{m_1-1}+s_{m_1-2})\cdots\rangle_s \\ &\times \langle \Theta(b+s_{m_2+1})\Theta(2b+s_{m_2+1}+s_{m_2+2}) \\ &\times \Theta(3b+s_{m_2+1}+s_{m_2+2}+s_{m_2+3})\cdots\rangle_s \end{aligned} \tag{A5}$$

and

$$P_{in}=\left\langle \Theta\left(-t-bM_b-\sum_{m=m_1+1}^{m_2} s_m\right)\Pi_L\Pi_R\right\rangle_s, \tag{A6}$$

where $\langle\cdots\rangle_s$ denotes the Gaussian average with probability density

$$f(s_m)=\frac{\exp\left(-\frac{1}{2}s_m^2\right)}{\sqrt{2\pi}}, \tag{A7}$$

and we introduced

$$\begin{aligned} \Pi_L\equiv&\Theta(-b-s_{m_1+1})\Theta(-2b-s_{m_1+1}-s_{m_1+2})\cdots \\ &\times\Theta(-bM_b-s_{m_1+1}-\cdots-s_{m_2}) \end{aligned} \tag{A8}$$

and

$$\begin{aligned} \Pi_R\equiv&\Theta(-b-s_{m_2})\Theta(-2b-s_{m_2}-s_{m_2-1})\cdots \\ &\times\Theta(-bM_b-s_{m_2}-\cdots-s_{m_1+1}). \end{aligned} \tag{A9}$$

The reason for the factorization of P into $P_{\rm out}$ and $P_{\rm in}$ is that a binding configuration $\{m_1,m_2\}$ which is energetically favorable compared to the binding configurations $\{m_1+n_1,m_2\}$ and $\{m_1,m_2+n_2\}$ with one of the end position fixed, is also energetically favorable compared to the configuration $\{m_1+n_1,m_2+n_2\}$ with both end positions changed.

Note that $P_{\rm out}$ itself also consists of two independent factors: the first factor excludes the configuration of binding regions with the end located to the left of m_1+1, while the second one excludes end positions larger than m_2. Both these factors can be written in the form

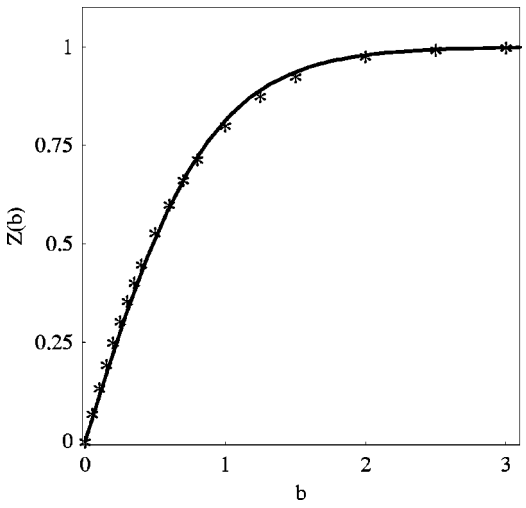

FIG. 9. Numerical solution of the function $Z(b)$ (stars). The best fit by a function of the form $Z(b)=\tanh(\kappa b)$ yields $\kappa\simeq 1.14$ (solid line).

$$Z(b)=\prod_{m=1}^{\infty}\left[\int_{-\infty}^{+\infty}ds_m f(s_m)\Theta\left(\sum_{k=1}^{m}(b+s_k)\right)\right], \tag{A10}$$

so that

$$P_{\text{out}}=[Z(b)]^2. \tag{A11}$$

We note that if the conditions imposed by the first Θ functions in Eq. (A10) are satisfied, the arguments of the last Θ functions also almost certainly are positive. In other words, only the first few Θ functions in Eq. (A10) are really restrictive and, therefore, we can replace the finite product of Θ functions by an infinite number of factors. To calculate Eq. (A10), we introduce the function $Z(s|b)$, satisfying the integral equation,

$$Z(s|b)=\int_0^{\infty}ds' f(s+b-s')Z(s'|b). \tag{A12}$$

Comparing the iterative solution of this equation to Eq. (A10), one finds $Z(b)=Z(0|b)$. The integral equation (A12) can be easily solved numerically. The result is shown as stars in Fig. 9 and the solid line represents the best fit to these points by a function of the form

$$Z(b)=\tanh(\kappa b) \tag{A13}$$

with $\kappa\simeq 1.14$.

The calculation of the inner factor is complicated by the presence of the extra Θ function in Eq. (A6), which precludes the factorization of P_{in} in two independent averages. However, considerable simplification is possible if we restrict our considerations to the average local binding energies with $1<bt<t$, where the main suppression factor in P is the probability of the local binding-energy fluctuation necessary to create a binding region at all (see Sec. III A). In other words, the most important contribution to P_{in} (and also P) comes from averaging the first Θ function in Eq. (A6). Bearing this in mind, we now calculate the inner factor Eq. (A6). First, we can rewrite Eq. (A6) in the form

$$P_{in}=\int_{t+bM_b}^{\infty}dS\int_{-i\infty}^{+i\infty}\frac{d\lambda}{2\pi i}e^{-\lambda S}\times\prod_{m=m_1+1}^{m_2}\int_{-\infty}^{+\infty}ds_m e^{-\lambda s_m}f(s_m)\Pi_L\Pi_R, \tag{A14}$$

where the integration over λ ensures that

$$S=-\sum_{m=m_1+1}^{m=m_2}s_m \tag{A15}$$

and the limits of the integration over S follow from the first Θ function in Eq. (A6). We now shift the argument of $f(s_m)$ on each site by λ,

$$f(s_m)\to f(s_m+\lambda)=e^{-(1/2)\lambda^2-\lambda s_m}f(s_m), \tag{A16}$$

so that the average value now becomes $s_m=-\lambda$ and Eq. (A14) reads

$$P_{\text{in}}=\int_{t+bM_b}^{\infty}dS\int_{-i\infty}^{+i\infty}\frac{d\lambda}{2\pi i}e^{-\lambda S+(1/2)M_b\lambda^2}\times\prod_{m=m_1+1}^{m_2}\int_{-\infty}^{+\infty}ds_m f(s_m+\lambda)\Pi_L\Pi_R. \tag{A17}$$

The integral over λ comes from the vicinity of $\lambda_0=S/M_b$, where the exponential in Eq. (A17) has its maximum. Saddle-point integration over λ then gives

$$P_{\text{in}}=\int_{t+bM_b}^{\infty}\frac{dS}{\sqrt{2\pi M_b}}e^{-(1/2M_b)S^2}\times\prod_{m=m_1+1}^{m_2}\int_{-\infty}^{+\infty}ds_m f\left(s_m+\frac{S}{M_b}\right)\Pi_L\Pi_R. \tag{A18}$$

Next we note that, as we saw before, only the first few Θ functions in Π_L and Π_R are really restrictive, because the relevant binding-energy fluctuation is close to the optimal fluctuation. This implies that the local binding-energy averages of Π_L and Π_R in Eq. (A18) are approximately decoupled. Furthermore, it is easily seen from Eq. (A10) that then $\langle\Pi_L\rangle=\langle\Pi_R\rangle=Z[(S/M_b)-b]$, so that Eq. (A18) becomes

$$P_{\text{in}}=\int_{t+bM_b}^{\infty}\frac{dS}{\sqrt{2\pi M_b}}e^{-(1/2M_b)S^2}\left[Z\left(\frac{S}{M_b}-b\right)\right]^2. \tag{A19}$$

The integral over S comes from the vicinity of the lower limit, $S=t+bM_b$. The result of the integration is

$$P_{\text{in}}=\sqrt{\frac{M_b}{2\pi}}\frac{\exp\left[-\frac{(t+bM_b)^2}{2M_b}\right]}{(t+bM_b)}\left[Z\left(\frac{a}{M_b}\right)\right]^2, \tag{A20}$$

where for S in the argument of the smooth function Z we took its value at the lower limit of the integration. Combining Eqs. (A1), (A11), and (A20), we obtain for the binding probability,

$$P=\frac{1}{2\mathcal{N}}\frac{\exp(-g^2)}{\sqrt{\pi}g}\left[Z\left(\frac{a}{M_b}\right)Z(b)\right]^2, \qquad \text{(A21)}$$

where $g=(t+bM_b)/\sqrt{2M_b}$ in accordance with the definition Eq. (15). The most important contribution to P is the exponential term in Eq. (A21), which is identified as the probability Eq. (16) of the binding energy fluctuation necessary to create binding between M_b pairs of contact residues at all (see Sec. III A). The last term in Eq. (A21) originates from the condition Eq. (24), which ensures that P is the binding probability of the optimal binding configuration with respect to all configurations that also contain a single binding region.

As the function Z is only slowly varying, we are able to calculate $\mathcal{N}$ from the normalization condition Eq. (12) by another saddle-point integration and obtain the final expression Eq. (25) for the binding probability in the original set of variables Eqs. (A2)–(A4).

[1] For a recent review, see C.E. Orsello, D.A. Lauffenburger, and D.A. Hammer, Trends Biotechnol. **19**, 310 (2001).

[2] For a general textbook on immunology, see I. Roitt, J. Brostoff, and D. Male, *Immunology*, 5th ed. (Mosby International, London, 2000).

[3] A.G. Amit, R.A. Mariuzza, S.E.V. Phillips, and R.J. Poljak, Science **233**, 747 (1986).

[4] S. Sheriff, E.W. Silverton, E.A. Padlan, G.H. Cohen, S.J. Smith-Gill, B.C. Finzel, and D.R. Davis, Proc. Natl. Acad. Sci. U.S.A. **84**, 8075 (1987).

[5] P.M. Colman, W.G. Laver, J.N. Varghese, A.T. Baker, P.A. Tulloch, G.M. Air, and R.G. Webster, Nature (London) **326**, 358 (1987).

[6] M. Cygler, D.R. Rose, and D.R. Bundle, Science **253**, 442 (1991).

[7] P. Ajitkumar, S.S. Geier, K.V. Kesari, F. Borriello, M. Nakagawa, J.A. Bluestone, M.A. Saper, D.C. Wiley, and S.G. Nathenson, Cell **54**, 47 (1988).

[8] J.K. Percus, O.E. Percus, and A.S. Perelson, Proc. Natl. Acad. Sci. U.S.A. **90**, 1691 (1993).

[9] A.S. Perelson and G. Weisbuch, Rev. Mod. Phys. **69**, 1219 (1997).

[10] J.K. Percus, O.E. Percus, and A.S. Perelson, J. Math. Biol. **40**, 278 (2000).

[11] The calculation of the binding probability P is in this case analogous to that of P_{in} [see Eq. (A6) in the Appendix] with the simplification $\Pi_L=\Pi_R=1$ or, equivalently, $Z=1$ in Eq. (A19).

[12] M.T. Figge, M. Mostovoy, and J. Knoester, Phys. Rev. B **58**, 2626 (1998).