PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Electrostatic rate enhancement and transient complex of protein–protein association

Ramzi Alsallaq[1,2,3] and Huan-Xiang Zhou[1,2,3]*

[1] Department of Physics, Florida State University, Tallahassee, Florida 32306
[2] Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306
[3] School of Computational Science, Florida State University, Tallahassee, Florida 32306

ABSTRACT

The association of two proteins is bounded by the rate at which they, via diffusion, find each other while in appropriate relative orientations. Orientational constraints restrict this rate to $\sim 10^5$–10^6 M^{-1} s^{-1}. Proteins with higher association rates generally have complementary electrostatic surfaces; proteins with lower association rates generally are slowed down by conformational changes upon complex formation. Previous studies (Zhou, Biophys J 1997;73:2441–2445) have shown that electrostatic enhancement of the diffusion-limited association rate can be accurately modeled by $k_D = k_{D0}\, exp(-\langle U_{el}\rangle^*/k_BT)$, where k_D and k_{D0} are the rates in the presence and absence of electrostatic interactions, respectively, $\langle U_{el}\rangle^*$ is the average electrostatic interaction energy in a "transient-complex" ensemble, and k_BT is the thermal energy. The transient-complex ensemble separates the bound state from the unbound state. Predictions of the transient-complex theory on four protein complexes were found to agree well with the experiment when the electrostatic interaction energy was calculated with the linearized Poisson–Boltzmann (PB) equation (Alsallaq and Zhou, Structure 2007;15:215–224). Here we show that the agreement is further improved when the nonlinear PB equation is used. These predictions are obtained with the dielectric boundary defined as the protein van der Waals surface. When the dielectric boundary is instead specified as the molecular surface, electrostatic interactions in the transient complex become repulsive and are thus predicted to retard association. Together these results demonstrate that the transient-complex theory is predictive of electrostatic rate enhancement and can help parameterize PB calculations.

Proteins 2008; 71:320–335.

***Key words:** protein association; electrostatic rate enhancement; diffusion control; transient complex; Poisson–Boltzmann equation.*

INTRODUCTION

The association of two proteins is bounded by the rate at which they find each other through diffusion. To form a stereospecific complex, the two molecules must have appropriate relative orientations when they come together. The orientational constraints severely restrict the diffusion-limited rate, which is estimated at 10^5–10^6 M^{-1} s^{-1} according to several theoretical models.[1–3] In particular, antibody–protein association rates are typically observed in this range.[4–6] However, in a wide array of biological processes, rapid association between proteins is crucial; in these cases, rates much higher than 10^6 M^{-1} s^{-1} have been observed. These include binding of cytotoxic nucleases with inhibitors (for self protection of the host cell),[7,8] binding of toxins to a potassium channel[9] and acetylcholinesterase[10] (for inhibition of neural transmission in the prey), and binding between many partners along the signaling pathway leading to the stimulation of actin polymerization.[11,12] In these cases the binding surfaces across the interface are usually found to have opposite electrostatic potentials (see Fig. 1), implicating electrostatic rate enhancement. Building on earlier work,[2,15–19] we have now fully developed a theory for predicting diffusion-limited association rates under the influence of electrostatic interactions.[20] The theory has been found to be very promising when tested on four protein complexes and 23 mutants over wide ranges of ionic strength.[21] Here we further apply this theory to examine electrostatic contributions to association rates.

The first theoretical result for the association rate was obtained by Smoluchowski,[22] who found that the diffusion-limited rate for two uniformly reacting spheres to form a complex is $k_{D0} = 4\pi DR$, where D is the relative translational diffusion constant and R is the contact distance between the two sphere centers. Debye recognized that the association rate between oppositely charged molecules can be increased by electrostatic interactions; for two uniformly reacting spheres with a centrosymmetric

Grant sponsor: NIH; Grant number: GM058187
*Correspondence to: Huan-Xiang Zhou, Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306. E-mail: zhou@sb.fsu.edu
Received 18 April 2007; Revised 12 June 2007; Accepted 15 June 2007
Published online 11 October 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21679

interaction potential $U(r)$, he found the diffusion-limited rate to be $k_D = 4\pi D / \int_R^\infty e^{U(r)/k_B T} r^2 \, dr$, where r is the intersphere distance and $k_B T$ is the thermal energy.[23] Throughout this paper we will use k_{D0} to denote the diffusion-limited rate in the absence of an interaction potential (also referred to as the basal rate), and use k_D to denote the counterpart in the presence of an interaction potential.

The stereospecific association of two proteins involves significant orientational constraints, and thus the formulas of Smoluchowski and Debye are of little use. Solc and Stockmayer tackled the problem of orientational constraints through a quasichemical approximation.[24] For two spheres each with a reactive patch, they found the diffusion-limited rate to be

$$k_{D0} = \frac{4\pi D R F_1 F_2}{\Lambda_1 \Lambda_2 + \left[(1-\Lambda_1)^{-1}(1-\Lambda_2)^{-1} + (1-\Lambda_1)^{-1}(\Lambda_2 - F_2)^{-1} + (1-\Lambda_2)^{-1}(\Lambda_1 - F_1)^{-1}\right]^{-1}} \quad (1)$$

where F_1 and F_2 are the surface fractions covered by the reactive patches. An approximate expression, obtained by Berg,[25] for Λ_i ($i = 1$ or 2), in the case of a patch spanning polar angles between 0 and δ_i, is given by

$$\Lambda_i / F_i = \frac{\xi_i + \cot(\delta_i/2)}{\xi_i + \sin(\delta_i/2)\cos(\delta_i/2)} \quad (2)$$

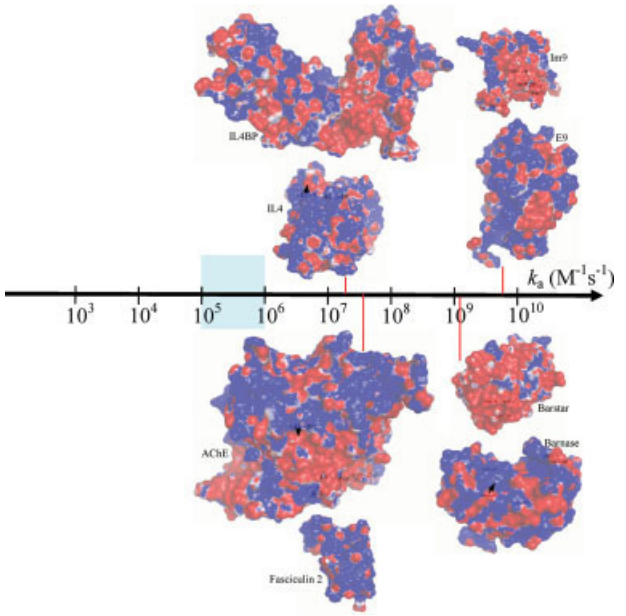

Figure 1

The wide spectrum of protein–protein association rates. The value 10^5 M^{-1} s^{-1} serves as the demarcation point separating the diffusion-limited regime from the conformational change-limited regime. In the diffusion-limited regime, rates in the narrow range of 10^5–10^6 M^{-1} s^{-1} are observed for association between proteins, such as antibody and antigen, which do not involve significant electrostatic contributions. Proteins that associate with higher rates typically have complementary electrostatic surfaces, as illustrated by the four protein complexes studied here. The association rates indicated for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes are experimental results measured at ionic strengths of 25, 13, 50, and 150 mM, respectively.[7,8,10,13] Electrostatic potential surfaces are generated by the APBS program[14] and displayed by PyMOL (http://www.pymol.sourceforge.net).

where $\xi_i = [(1 + D_i R^2/D)/2]^{1/2}$ and D_i is the rotational diffusion constant. For small patches one finds

$$k_{D0}/4\pi DR = F_1 \xi_2 \tan(\delta_2/2) + F_2 \xi_1 \tan(\delta_1/2) \quad (3)$$

For medium-sized proteins, R, D, and D_i are of the orders of 40 Å, 20 $Å^2$ ns^{-1}, and 0.02 ns^{-1}, respectively. The basal rate for forming a complex with three or more stereospecific contacts, obtained by Brownian dynamics simulations,[1] is 10^5–10^6 M^{-1} s^{-1}, which is four orders of magnitude lower than the unrealistic Smoluchowski result. Such a rate is predicted by Eq. (3) for reactive patches with $\delta_1 \sim \delta_2 \sim 5^\circ$. Results similar to Eq. (1) for the basal rate have been obtained by a number of different methods.[15,26–28] Within the model of two spheres with reactive patches, the influence of an interaction potential on the association rate has also been studied and expressions for k_D in the presence of a centrosymmetric potential were obtained.[2,15]

Brownian dynamics simulations make it possible to calculate the association rate for protein shapes and interaction potentials beyond the scope of analytical theories.[1,15,29–37] In one Brownian dynamics simulation study,[15] it was discovered that the rate enhanced by electrostatic interactions to a good approximation is given by

$$k_D = k_{D0} \exp(-\langle U_{el} \rangle^* / k_B T) \quad (4)$$

where $\langle U_{el} \rangle^*$ is the average electrostatic interaction energy of the proteins in the reactive region. Analysis on the expressions of k_D for the model of two spheres with reactive patches in the presence of an interaction potential showed that the validity of Eq. (4) depends on two conditions.[2,15] First, the reaction region is small, meaning that the complex formed is stereospecific. Second, the interaction potential is long-ranged. A proof of Eq. (4) as a good approximation under the two conditions was also found.[16] The necessity for the condition of small reaction region can be demonstrated on the model of two uniformly reacting spheres. A blind use of Eq. (4) to this model would predict $k_D = 4\pi DR \exp[-U(R)]/k_B T]$,

which significantly overestimates the effect of the interaction potential, as found in the exact result of Debye.

Fortuitously (and fortunately) the two conditions for the accuracy of Eq. (4) are satisfied for the stereospecific association of proteins under the influence of electrostatic interactions. The accuracy of Eq. (4) has been demonstrated on protein–ligand binding rates by direct comparison with results for k_D obtained by Brownian dynamics simulations.[17,38,39] Equation (4) allows the diffusion-limited rate to be calculated without expensive Brownian dynamics simulations in the presence of electrostatic interactions. The computational cost of calculating forces and torques in such simulations has necessitated simplified treatment of electrostatic interactions.[29–32]

The overall association process between two proteins, A and B, can be viewed as consisting of two steps:

$$A + B \underset{k_{-D}}{\overset{k_D}{\rightleftharpoons}} A^*B \xrightarrow{k_c} AB \tag{5}$$

The first is to reach a reactive region (with rate k_D) through translational and rotational diffusion; the transient complex thus formed, A*B, can also dissociate (with rate k_{-D}). The second is to reach the final bound state (with rate k_c) through conformational rearrangements. Assuming that the transient complex is in steady state, the overall association rate is

$$k_a = \frac{k_D k_c}{k_{-D} + k_c} \tag{6}$$

The ensemble of configurations making up the transient complex serves as the dividing surface between the bound and unbound states. Something like this ensemble has been implied by a loosely defined term called "encounter complex."[30] As demonstrated on a model system,[40] the overall association rate k_a calculated according to Eq. (6) should be insensitive to the precise location of the dividing surface. However, practical considerations lead to an unequivocal compromise for the specification of the transient complex.[19,20] In general, overall translational and rotational diffusion and long-range electrostatic interactions dominate outside the dividing surface, whereas conformational rearrangements and short-range interactions become dominant inside the dividing surface. For the diffusion-limited rate k_D to be a good estimate for the overall association rate k_a, the dividing surface should be as close to the bound state as possible. On the other hand, placing the dividing surface inside the potential well of the bound state would require modeling conformational rearrangements along with overall translational and rotational diffusion. The compromise is then to put the dividing surface right at the outer boundary of the bound-state potential well.[20]

Note that the ensemble of configurations on the dividing surface, formerly referred to as the transition state for association,[18–21] is now referred to as the transient complex. The change in nomenclature is intended to avoid misidentification of Eq. (4) as a naive application of the Eyring-type transition-state theory developed for barrier crossing.[41] Equation (4) is fundamentally different from Eyring's transition-state theory. First of all, the former is applied to a diffusion-limited process—reaching the transient complex does not require energy activation, whereas the latter is applied to an activation-limited process. Second, k_D only involves the process up to the formation of the transient complex; what happens afterward has no bearing. Effectively the transient-complex ensemble serves as an absorbing boundary. In contrast, the speed of decomposition of the activated complex is an integral part of the formulation of Eyring's transition-state theory. Third, the prefactor k_{D0} is the rate in the absence of interactions; favorable interactions, corresponding to a negative $\langle U_{el}\rangle^*$, in the transient complex serves to enhance the association rate, opposite to the role played by an activation barrier in Eyring's theory. Fourth, unlike Eyring's theory, Eq. (4) does not require a reaction coordinate. In fact, as we will describe later, the specification of the transient-complex ensemble requires at least six degrees of freedom—three for relative translation and three for relative rotation. In short, there is no connection between Eq. (4) and Eyring's transition-state theory.

Equation (4) together with the specification of the transient-complex ensemble will now be referred to as the transient-complex theory for protein association. This theory was tested on a large set of experimental data for association rates, with promising results.[21] In this paper we use the theory and the experimental data to investigate details of calculating electrostatic contributions to association rates. We find that calculations with the nonlinear Poisson–Boltzmann (PB) equation instead of the linearized version lead to an improvement, albeit modest, in agreement between theory and experiment. In addition, changing the boundary between the protein and solvent dielectrics from the van der Waals (vdW) surface to the molecular surface (MS) turns electrostatic rate enhancement into retardation.

THEORETICAL METHODS

Protein complexes studied

The same four protein complexes studied in an earlier paper[21] are studied here. They are formed between colicin E9 and immunity protein Im9, between barnase (Bn) and barstar (Bs), between acetylcholinesterase (AChE) and fasciculin 2 (Fas), and between interleukin-4 (IL4) and IL4-binding protein (IL4BP). The protein data bank codes for the X-ray structures of the native

Table I
Overall Properties of Four Protein Complexes

Complex	PDB[a]	No. of atoms[b]	Net charges[b]	No. of interface atoms	N_c in native complex[c]	N_c^*
E9:Im9	1emvB; A	2124; 1308	+7; −9	203	42 (18)	24
Bn:Bs	1brsC; F	1727; 1433	+2; −6	210	38 (21)	14
AChE:Fas	1mahA; F	8340; 906	−9; +4	271	52 (23)	23
IL4:IL4BP	1iarA; B	2111; 2959	+7; −5	181	35 (14)	16

[a]X-ray structures were determined by Kuhlmann *et al.*,[42] Buckle *et al.*,[43] Bourne *et al.*,[44] and Hage *et al.*,[45] respectively. After the PDB entry name, the chains (separated by a semicolon) used for the two subunits in each complex are also listed.
[b]Data for the two subunits in each complex are given separately.
[c]The number of native contacts among all contacts are given inside parentheses.

complexes, and the number of atoms and net charge of each subunit are listed in Table I. We will use the abbreviated names of two proteins, separated by a colon, in the form of E9:Im9, to denote a complex.

Our study also covers 23 mutants of the four protein complexes, for which experimental data on association rates are available.[8,10,13,46–49] There are five single mutants on the E9:Im9 complex, each replacing an Im9 residue by alanine. These are E9:E30A; E9:E41A; E9:S50A; E9:D51A; and E9:Y55A. Note that, throughout the paper, substitutions, such as E30A, before and after a colon refer to mutations on the first and second protein, respectively, of a complex. Twelve of the mutants are on the Bn:Bs complexes: K27A:Bs; R59A:Bs; E60A:Bs; R83Q:Bs; R87A:Bs; Bn:D35A; Bn:D39A; Bn:E76A; K27A:D39A; R59A:D35A; R83Q:D39A; and R87A:D39A. There are four AChE:Fas mutants (Fas:D74N; Fas:E202Q; Fas:D280V; and Fas:D283Q) and two IL4:IL4BP mutants (E9Q:IL4BP and R88A:IL4BP). All mutations, except for E202Q:Fas, are in or around the interfaces of the complexes. Two mutations, E60A:Bs and E9Q:IL4BP, increase the net charges on the mutated proteins; the E9:Y55A reduce the polarity of the mutated side chain; all the other mutations decrease the net charges on the mutated proteins.

As described previously,[50,51] hydrogens are added to all heavy atoms in the X-ray structure of each protein complex and energy-minimized in the InsightII program (Accelrys, San Diego). Each mutation is modeled by replacing the side chain and energy-minimizing its conformation with the rest of the protein complex fixed.

Energy landscape for protein association

As noted in the Introduction, the transient complex on the pathway to stereospecific association consists of the ensemble of configurations located at the outer boundary of the bound-state potential well. The identification of these configurations, as outlined previously,[20,21] is based on mapping the energy landscape over the bound state and the surrounding region. To that end, configurations in the bound state and the transition region to the unbound state are uniformly sampled. The two subunits in a protein complex are treated as rigid; therefore, there are only six relevant degrees of freedom: three for relative translation and three for relative rotation [Fig. 2(a)]. Each subunit is frozen in its conformation found in the native complex, leading to a smoothed energy landscape. For easy reference, the larger and smaller subunits will be designated A and B, respectively.

The six relative translational and rotational coordinates are defined as follows. On the X-ray structure of the native complex, interface atoms, taken as heavy atoms having interfacial contacts less than 5 Å, are collected (the number of interface atoms in each of the four protein complexes studied is listed in Table I). The geometric center of the interface atoms is body-fixed on the two subunits, to become the centers of the respective binding surfaces. The normal to the least-squares plane of the interface atoms, pointing from subunit A to subunit B, is also body-fixed on the two subunits to define two unit vectors. Subunit A is then fixed in the laboratory frame, and subunit B is translated and rotated. The vector, **r**, from the center of the binding surface on subunit A to the counterpart on subunit B defines the relative translational coordinates. The unit vector, **e**, attached to subunit B and the rotation angle, χ, around the unit vector together define the three relative rotational coordinates.

The displacement vector **r** is represented in spherical coordinates (r, θ, ϕ), with the laboratory z axis along the unit vector attached to subunit A. Subunit B has a body-fixed frame, with z axis along the unit vector **e**. The rotation of subunit B is represented by the three Euler angles (ξ, ζ, χ) of the body-fixed frame relative to the laboratory frame. Note that ξ and ζ are the polar and azimuthal angles, respectively, of the unit vector **e** in the laboratory frame. The native complex corresponds to $r = 0$, $\xi = 0$, and $\chi = 0$. To sample the bound state and the transition region to the unbound state, the six translational and rotational coordinates are randomly generated, with only one restriction: $r \leq r_0$. The value of r_0 is set to 6 Å.

The energy landscape over the bound state and the transition region to the unbound state is mapped by

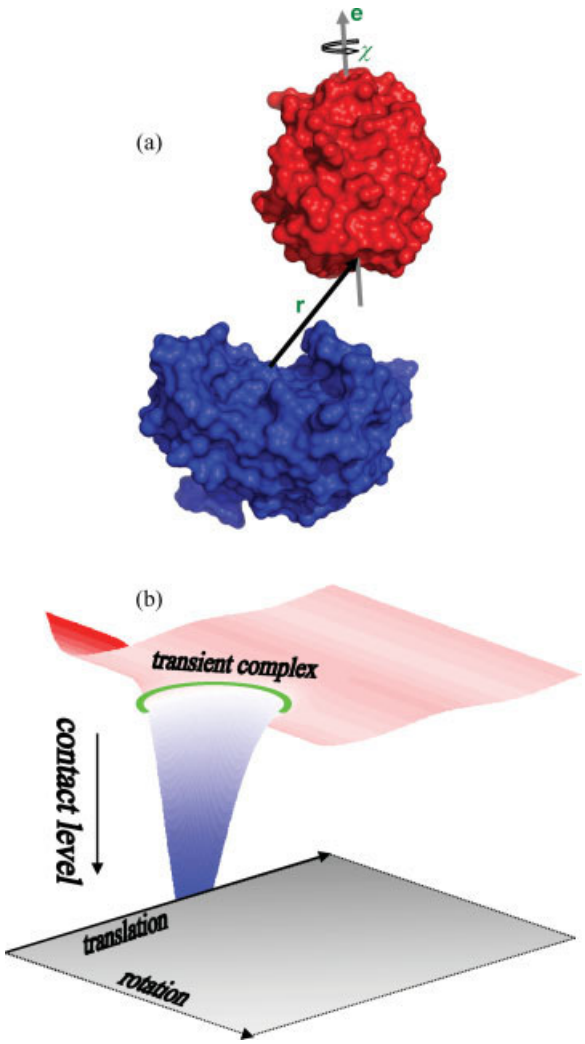

Figure 2

*(**a**) Definition of six translational and rotational coordinates for two associating proteins. One protein, shown in blue, is fixed in space; the other, shown in red, can freely translate and rotate. The three translational degrees of freedom are represented by the displacement vector **r** between the centers of the binding surfaces on the two proteins. Of the three rotational degrees of freedom, two are a unit vector **e** attached to the moving protein and the remaining one is the rotational angle χ around the unit vector. The unit vector is perpendicular to a plane defined by the binding surface. (**b**) Illustration of the energy landscape for protein–protein association. The bound state is located in a deep "well" with high levels of contact. The transient-complex ensemble, indicated by a green ring, marks the termination of sharp decrease in contact level and the onset of sharp increase in translational and rotation freedom. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

detecting steric clashes, which are to be avoided, and by calculating the number of contacts, which favor the bound state, in the randomly generated configurations. Clash is defined through clash distances assigned to three types of atoms: hydrogen, polar (nitrogen and oxygen), and nonpolar (carbon and others). The clash distance within one type or between two types of atoms is set to the minimum distance of such pairs in the X-ray structure of the native complex. Typical values are 2.5–2.7 Å between polar atoms, 3.2–3.5 Å between nonpolar atoms, 2.8–3.1 Å between polar and nonpolar pairs, 1.6–2.1 Å between hydrogens, 1.6–1.7 Å between polar and hydrogen atoms, and ∼2.5 Å between nonpolar and hydrogen atoms. Exhaustive clash detection involves testing all atoms on subunit A against all atoms on subunit B. To speed up this process, atoms are trimmed from the two full lists as much as possible before testing for clash. First, the minimum and maximum z coordinates of each subunit are found; these, expanded from both below and above by the maximum clash distance, define the range of z coordinates for the subunit. The overlap region of the z ranges of the two subunits is then obtained. If the overlap is empty, the two subunits cannot clash. Otherwise atoms on the two subunits, with z coordinates within the overlap region, are retained for further test. This procedure is repeated for x and y coordinates, each time starting from the previous trimmed lists of atoms. The two final trimmed lists of atoms are tested exhaustively for clash.

For configurations that pass the clash test, contacts, either native or nonnative, are counted. Contacts are assumed to occur between interaction-locus atoms, which are selected from the interface atoms. The purpose of the selection is twofold: to increase the chance that retained native contacts are distinct from each other; and to decrease the chance of nonnative contacts so that there is a proper balance between native and nonnative contacts. The selection for interaction-locus atoms proceeds as follows. All cross-interface pairs are sorted in ascending order of interatomic distances in the native complex; each pair is evaluated against preceding pairs for possible elimination. Specifically, a pair is eliminated if it is within 3.5 Å of a preceding pair on either side of the interface in the native complex. The final remaining list constitutes the native pairs of interaction-locus atoms. The number of native pairs of interaction-locus atoms in each of the four protein complexes studied is listed in Table I.

Now the identification of native and nonnative contacts in clash-free configurations can be described. For that purpose, the interatomic distance of each native pair is equally divided between the two interaction-locus atoms across the interface to define their contact radii. For a native pair of interaction-locus atoms, the upper limit in distance for forming a contact is the native contact distance (equal to the sum of their contact radii) plus 3.5 Å. For a nonnative pair of interaction-locus atoms to form a contact, their distance has to be within the upper limit set at the sum of their contact radii plus 2.5 Å. The total number of contacts, both native and nonnative, is recorded for each configuration. This will be denoted as N_c. The value of N_c for each of the four protein complexes studied, in its X-ray structure, is also listed in Table I.

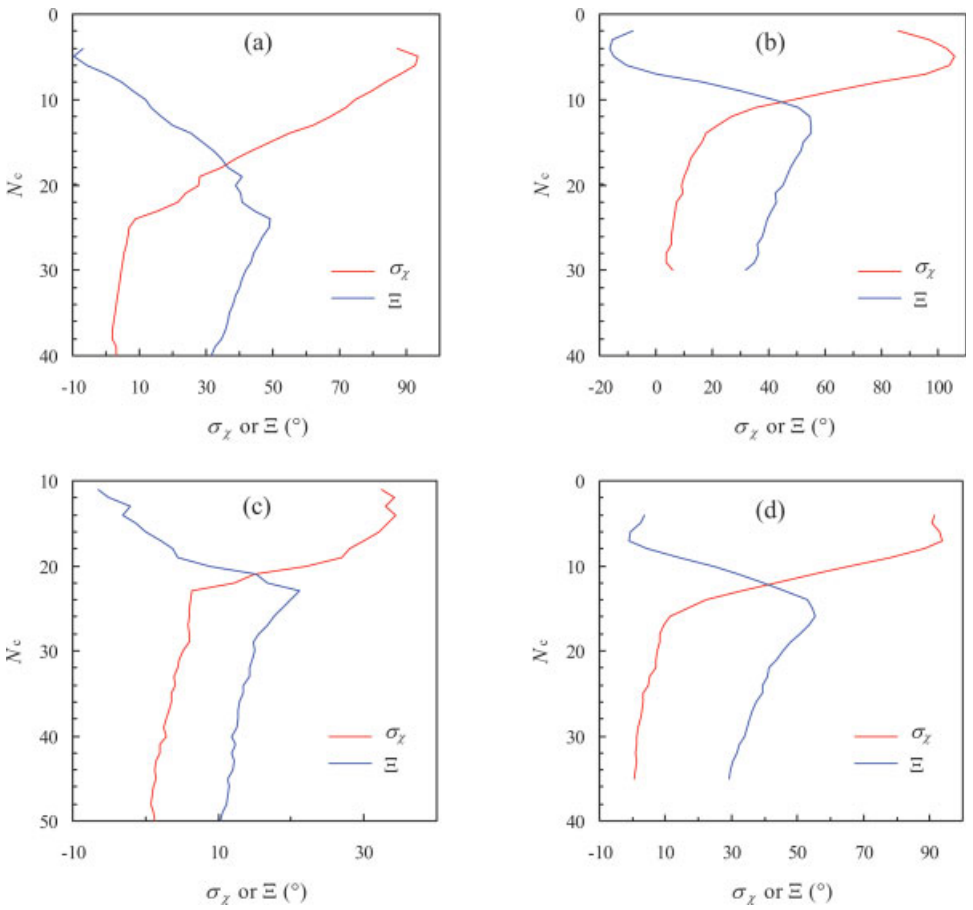

Figure 3

*Transition of the standard deviation of χ, σ_χ, from the bound state (with high contact levels) to the unbound state (with low contact levels). The start of the sharp increase in σ_χ marks the transient complex, with the corresponding contact level N_c^* uniquely determined by the maximum of Ξ. Panels (**a**)–(**d**) are for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

The energy landscape over the bound state and the transition region to the unbound state, as represented by the contact level N_c as a function of the six translational and rotational coordinates, has general features illustrated in Figure 2(b).[20] The bound state is represented by a deep well, in which translation and rotation are restricted. Outside the deep well, at most a few contacts are made, but translational and rotational freedom is gained. The outer boundary of the bound state marks the end point of a sharp decrease in contact level and the start point of a sharp increase in translational and rotational freedom.

Specification of the transient complex

The energy landscape for stereospecific association, featuring a sharp transition from contact domination in the bound state to translational/rotational-freedom domination in the unbound state, allows for an unequivocal specification of the transient complex. We identify the transient complex with the contact level, N_c^*, at which the transition occurs. In particular, when $N_c > N_c^*$, the sampled χ values are restricted to a small range around zero. As N_c further decreases, the range of sampled χ values sharply increases. Therefore $\sigma_\chi(N_c)$, the standard deviation of χ sampled at a given contact level, experiences a sharp increase at N_c^* (see Fig. 3). The sharp increase in σ_χ coincides with the maximum in the difference, Ξ, between σ_χ and its average at all lower contact levels. We thus obtain N_c^* as the contact level at which the function

$$\Xi(N_c) = \langle \sigma_\chi(N_c') \rangle_{N_c' < N_c} - \sigma_\chi(N_c) \qquad (7)$$

is maximal (see Fig. 3). In all, 436,058, 1,033,047, 74,620, and 231,683 collision-free configurations are accumulated for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes, respectively. The transient-complex contact levels identified from these configurations are listed in Table I.

Configurational volume of the bound state

The bound state is now explicitly defined by the condition $N_c > N_c^*$. The configurational volume of the bound-state region is given by

$$\mathcal{V}_b = (8\pi^2)^{-1} \times \int_{N_c > N_c^*} I(r, \theta, \phi, \xi, \zeta, \chi) r^2 \, dr \, d\cos\theta \, d\phi \, d\cos\xi \, d\zeta \quad (8)$$

where I is 1 if the configuration (r, θ, ϕ, ξ, ζ, χ) is clash-free and 0 otherwise. In our configurational sampling, r is randomly picked from 0 to r_0, cos θ and cos ξ are randomly picked from -1 to 1, and ϕ, ζ, and χ are randomly picked from 0 to 2π. Equation (8) can be expressed as

$$\mathcal{V}_b = 4\pi f_c f_b r_0 \langle r^2 \rangle_b \quad (9)$$

where f_c is the fraction of configurations that are clash-free among all the randomly generated configurations, f_b is the fraction of configurations that satisfy $N_c > N_c^*$ among all clash-free configurations, and $\langle r^2 \rangle_b$ is the average of r^2 among all bound-state configurations. In particular, as described later, $\mathcal{V}_b$ will play a prominent role in the calculation of the basal association rate k_{D0}.

Calculation of the electrostatic interaction free energy

For each of the four protein complexes studied, the transient-complex ensemble is represented by 100 configurations with the contact level N_c^*. The electrostatic interaction energy in each configuration is calculated as

$$U_{el} = U_{el}(AB) - U_{el}(A) - U_{el}(B) \quad (10)$$

The three terms on the right-hand side represent the electrostatic energies of the protein complex and the two subunits each by itself, respectively. The same 100 configurations of a wild-type complex are used for all its mutants. To that end, the two subunits in the native complex of each mutant are superimposed to their counterparts in each of the 100 transient-complex configurations. The average electrostatic interaction energy, $\langle U_{el} \rangle^*$, of the transient complex is calculated over the 100 configurations.

Electrostatic energies are calculated by solving the PB equation using the UHBD program.[52] Four types of calculations are carried out. Either the nonlinear or linearized PB equation is solved, with the dielectric boundary specified to be either the vdW surface or the MS. The latter is defined as the surface of the solute region excluded to a 1.4-Å solvent probe, and also known as the solvent-exclusion surface. The nonlinear PB equation is selected by adding the "full" option in the input script for the UHBD program; the MS specification is selected by adding the "nmap 1.5, nsph 500" option. Other details of the UHBD calculations are described in previous studies of salt and mutation effects on the binding stability of the four protein complexes.[50,51]

Protein charges are taken from the Amber force field,[53] and atomic radii are adapted from OPLS[54] and Bondi radii,[55] with values of 1.9, 1.2, 1.625, 1.48, and 1.775 Å for C, H, N, O, and S, respectively. The temperature is 298 K; the solute and solvent dielectric constants are 4 and 78.5, respectively. Ionic strengths correspond to salt concentrations in the experimental studies.[7,8,10,13,46–49]

Determination of basal rate by Brownian dynamics simulations

The basal rate k_{D0} in Eq. (4) refers to the diffusion-limited rate for reaching the transient complex in the absence of electrostatic interactions between the two subunits. This can be obtained by simulating the translational and rotational motion of the proteins. For this purpose we follow an algorithm developed previously.[15,56]

To start a Brownian dynamics simulation, subunit A is fixed in the laboratory frame, and subunit B is randomly placed in the bound state, which is specified by the condition $N_c > N_c^*$. While inside the bound state, subunit B is assigned a uniform rate γ for reacting with subunit A to form the native complex. A trajectory consists of repetitions of three steps. (1) A move of force-free Brownian translation and rotation is attempted. (2) Steric clash with subunit A is tested on the trial move. If the trial move is clash-free, it is accepted; otherwise the old configuration is accepted. (3) Whether reaction with subunit A has occurred during this move is tested. If yes, the trajectory is terminated. All trajectories are otherwise propagated to a preset cutoff time t_{cut}. For each trajectory, the lifetime of the trajectory is recorded for later use. First let us further explain the three steps.

For generating the Brownian translation and rotation of subunit B, we follow the extension of the Ermak–McCammon algorithm[57] by Fernandes and de la Torre.[58] For simplicity, we set the center of diffusion to the center of geometry, and assume isotropic translational and rotational diffusion. The translational and rotational diffusion constants are denoted as D and D_r, respectively. The principal axes of rotation are aligned with the axes of the body-fixed frame in subunit B. The center of diffusion, $\mathbf{R}$, is propagated according to

$$\mathbf{R}(t_1) = \mathbf{R}(t_0) + \boldsymbol{\mathcal{M}}^{b \to l}(t_0) \cdot \Delta\mathbf{s} \quad (11)$$

where $t_1 = t_0 + \Delta t$, $\boldsymbol{\mathcal{M}}^{b \to l}$ is the rotation matrix for transforming a body-fixed vector into the laboratory frame, and $\Delta\mathbf{s}$ is a displacement vector with components sampled from a Gaussian distribution, which has a

standard deviation of $(2D\Delta t)^{1/2}$. If $\Delta\mathbf{s}$ is replaced by the displacement vector, in the body-fixed frame, from the center of diffusion to the center of the binding surface, Eq. (11) would give the coordinates of the center of the binding surface on subunit B (i.e., $\mathbf{r}$) in the laboratory frame. The rotation matrix $\mathcal{M}^{b\to l}$ is propagated by rotations around the three principal axes,

$$\mathcal{M}^{b\to l}(t_1) = \mathcal{M}^{b\to l}(t_0)\cdot\mathcal{B}_x(\Delta\phi_x)\cdot\mathcal{B}_y(\Delta\phi_y)\cdot\mathcal{B}_z(\Delta\phi_z) \quad (12)$$

where $\mathcal{B}_{x,y,z}$ are the usual transformation matrices for rotations around the principal axes. The rotational angles $\Delta\phi_{x,y,z}$ are sampled from a Gaussian distribution with a standard deviation of $(2D_r\Delta t)^{1/2}$. The relative translational diffusion constant, D, is the sum of diffusion constants of the two subunits. Diffusion constants are estimated from the molecular weights of the subunits.[59] The resulting translational diffusion constants are 23, 24, 19, and 21 $\text{Å}^2\ \text{ns}^{-1}$, respectively, for E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP. The rotational diffusion constants of the moving subunits in these complexes, Im9, Bs, Fas, and IL4BP, are 0.029, 0.028, 0.035, and 0.015, respectively. Variable timesteps are used. When $r < r_0$, Δt is set to a constant value $\sim 5\times 10^{-5}$ ns; when $r > r_0$ Å, Δt is increased by $10^{-2}(r - r_0)^2/2D$. The value of r_0 is 6 Å.

Test for steric clash is done in the same way as described earlier for mapping the energy landscape. A minor exception is that, when subunit B is too far away, clash is not possible and thus not tested. Also in keeping with configurational sampling done for mapping the energy landscape, subunit A is fixed in the laboratory frame. This is done to speed up the test for steric clash, which is the most expensive part of the simulation. The neglect of rotational diffusion for subunit A leads to a very small underestimate of the diffusion-limited association rate.

Whether reaction occurs during a move is determined by comparing the probability for reaction with a random number uniformly distributed between 0 and 1. This reaction probability is $\exp(-\gamma\Delta t)$ if subunit B is inside the bound-state region both before and after the move, is $\exp(-\gamma\Delta t/2)$ if subunit B moves into or moves out of the bound-state region by the move, and is 0 if subunit B stays in the unbound region. Reaction occurs when the random number is smaller than the reaction probability.

The bound-state configurations, i.e., those with $N_c > N_c^*$, obtained in the sampling for mapping the energy landscape are used as initial configurations for the Brownian dynamics simulations. For the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes, 24,174, 45,046, 14,711, and 5130 trajectories, respectively, are generated. From the lifetimes of the trajectories, the survival fraction, $S(t)$, at times up to t_{cut} is calculated. This survival fraction is equal to the time-dependent association rate coefficient, $k_a(t)$, scaled by its initial value, $k_a(0) = \mathcal{V}_b\gamma$.[60] We thus have

$$k_a(t) = \mathcal{V}_b\gamma S(t) \quad (13)$$

The steady-state value, $k_a(\infty)$, is the desired association rate constant. This value is obtained by extrapolation through the asymptotic behavior of $k_a(t)$,[56]

$$k_a(t) = k_a(\infty)\left[1 + k_a(\infty)(\pi Dt)^{-1/2}/4\pi D + \cdots\right] \quad (14)$$

Specifically, $k_a(\infty)$ is obtained as the intercept of a fit of the long-time portion of $k_a(t)$ as a linear function of $(\pi Dt)^{-1/2}$. The linearity is ensured when t_{cut} is sufficiently large. The value of t_{cut} used is $\sim 10^4$ ns. The association rate thus obtained depends on the reactivity γ. The actual value of γ, 25–50 ns^{-1}, is selected to make $S(t)$ approach 0.5 at $t = t_{cut}$, such that the uncertainty on $S(t_{cut})$ is minimized. The diffusion-limited association rate k_{D0} is the limit of $k_a(\infty)$ at $\gamma = \infty$. This is obtained from the formula[56]

$$\frac{1}{k_a(\infty)} = \frac{1}{k(0)} + \frac{1}{k_{D0}} \quad (15)$$

In comparing association rates predicted by Eq. (4) against experimental data, we treat the basal rate k_{D0} as an adjustable parameter. Specifically, for each protein complex, the value of k_{D0} is obtained by fitting experimental data for the salt dependence of the association rate of the wild-type proteins. This fitted value is then used without further adjustment for all the mutants. Whether any discrepancy between the fitted value and the result from the Brownian dynamics simulations is acceptable is addressed later.

RESULTS

Placement of the transient complex

An essential part of our theory for predicting electrostatic enhancement of protein association rates is the specification of the transient complex. In general terms, the transient complex separates the bound state, with numerous short-range interactions but restricted translational and rotational freedom, from the unbound state, with at most a small number of interactions but expanded configurational freedom. As reasoned in the Introduction, we specifically place the transient complex at the outer boundary of the bound state. As Figure 3 shows, the outer boundary of the bound state corresponds to the onset of a sharp increase in the sampling range of the rotational angle χ, as measured by the standard deviation σ_χ. The contact level, N_c^*, defining the transient complex is uniquely identified by the maximum in the function Ξ given in Eq. (7). The transient-complex contact levels are 24, 14, 23, and 16,

respectively, for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes. These are to be compared with the corresponding contact levels of 42, 38, 52, and 35 in the native complexes.

The full energy landscape over the bound state, the transient complex, and beyond can be presented by a seven-dimensional scatter plot, consisting of points that pair configurations with corresponding contact levels. Glimpses into this energy landscape are provided by projections of the seven-dimensional scatter plot into the N_c-χ and N_c-r planes, as shown in Figure 4. These projected scatter plots clearly show the sharp transition from contact domination in the bound state to translational/rotational-freedom domination in the unbound state. (The transition in the N_c-r plots could be better shown had we extended the upper limit in r from 6 Å during the random sampling of configurations, as was presented previously for the Bn-Bs complex with a upper limit of r at 10 Å.[20] The choice of the 6-Å upper limit is made to ensure adequate sampling of the bound state and the transient complex.) In the N_c-χ and N_c-r plots shown in Figure 4, the transient-complex contact level, N_c^*, is also highlighted. As expected, this contact level is precisely where the transition occurs. For the other four coordinates, θ, ϕ, ξ, ζ, there are no apparent transitions in the ranges of sampled values (data not shown).

In the N_c-χ and N_c-r plots, absence of configurations that have high contact levels but large rotations or large separations and configurations that have small separations but high contact levels are reflections of the deep potential well of the bound state [as illustrated in Fig. 2(b)]. However, there are also voids around $\chi = \pm 90^\circ$, regardless of contact level, for the E9:Im9 and AChE:Fas complexes, and at χ around $+90^\circ$ for the IL4:IL4BP complex. The voids arise from the shape complementarity of the binding surfaces across the interface of a protein complex. Rotating one subunit by 90° with respect to the other subunit may lead to steric clash, which eliminates that configuration in the N_c-χ plot. There is an apparent asymmetry between $\chi = +90^\circ$ and $\chi = -90^\circ$ in the IL4:IL4BP complex. Interestingly, in all the four protein complexes, rotation angles around 180° can be sampled with moderately high contact levels, indicating that a 180° flip can be tolerated more than a 90° rotation. This is in line with the well-known difficulty of outscoring the correct configuration against a 180°-flipped alternative in protein–protein docking.

Among the transient-complex configurations of the E9:Im9, Bn:Bs, and AChE:Fas complexes, positive and negative values of χ are sampled nearly equally, leading to averages of χ close to zero. In the IL4:IL4BP complex, however, there is a bias toward negative χ, consistent with the aforementioned asymmetry between $\chi = +90^\circ$ and $\chi = -90^\circ$; the average χ value in the transient complex is -9°. The standard deviations of transient-complex χ are 9, 18, 6, and 11°, respectively, for the four protein complexes (see Fig. 3). The variation among the four complexes, with Bn:Bs showing the largest spread in χ, appears to be dictated by the overall curvatures and smoothness of the binding surfaces. The relative separation r averages 3.9, 4.9, 5.1, and 4.4 Å, respectively, among the transient-complex configurations of the four protein complexes, with standard deviations all around 0.5 Å.

Basal rate from Brownian dynamics simulations

The basal rate k_{D0} is calculated from the initial value $k_a(0) = \mathcal{V}_b\gamma$ and the long-time limit, $k_a(\infty)$, of the time-dependent association rate $k_a(t)$, according to Eq. (15). To obtain the bound-state configurational volume $\mathcal{V}_b$, the clash-free fraction f_c, the bound fraction f_b, and $\langle r^2\rangle_b$ are needed [see Eq. (9)]. The values of f_c are 5.4×10^{-5}, 1.0×10^{-4}, 9.0×10^{-6}, and 8.6×10^{-5}, respectively, for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes. The corresponding values of f_b are 0.059, 0.055, 0.198, and 0.032, and the corresponding values of $\langle r^2\rangle_b$ are 19.5, 10.5, 20.3, and 13.4 Å^2. The resulting values of $\mathcal{V}_b$ are 2.4×10^{-3}, 8.7×10^{-3}, 2.7×10^{-3}, and 2.8×10^{-3} Å^3 for the four protein complexes.

With $\gamma = 26$, 26, 50, and 26 ns^{-1} for the four complexes, $k_a(0)$ has values of 0.06, 0.23, 0.14, and 0.07 Å^3 ns^{-1}, and the corresponding values of $k_a(\infty)$ from the Brownian dynamics simulations are 0.03, 0.13, 0.04, and 0.04 Å^3 ns^{-1}. The basal rates are thus 3.4×10^4, 1.9×10^5, 3.6×10^4, and 4.1×10^4 M^{-1} s^{-1} for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes. The higher k_{D0} value of the Bn:Bs complex appears to be a result of the wider span of its transient complex in configurational space, as indicated by the fact that Bn:Bs has the largest transient-complex σ_χ and the largest $\mathcal{V}_b$ among the four protein complexes.

As will be seen shortly, the basal rates obtained from the Brownian dynamics simulations underestimate the values required to fit experimental data for the salt dependences of the association rates of the protein complexes, by up to an order of magnitude. As noted previously,[21] the underestimate is not unexpected, since in the Brownian dynamics simulations the proteins are treated as rigid. Local induced fit would allow the two proteins to reach configurations that are forbidden in the rigid treatment, thus enlarging the span of the transient-complex configurations and possibly leading to the desired higher basal rate. An additional explanation for the discrepancy on the basal rates will be presented under Discussion.

Salt dependences of association rates for wild-type complexes

In our previous study,[21] we have already found that the transient-complex theory predicts well the salt and mutation effects on the association rates of the four protein complexes when the average electrostatic interaction energy, $\langle U_{el}\rangle^*$, is calculated by the linearized PB equation

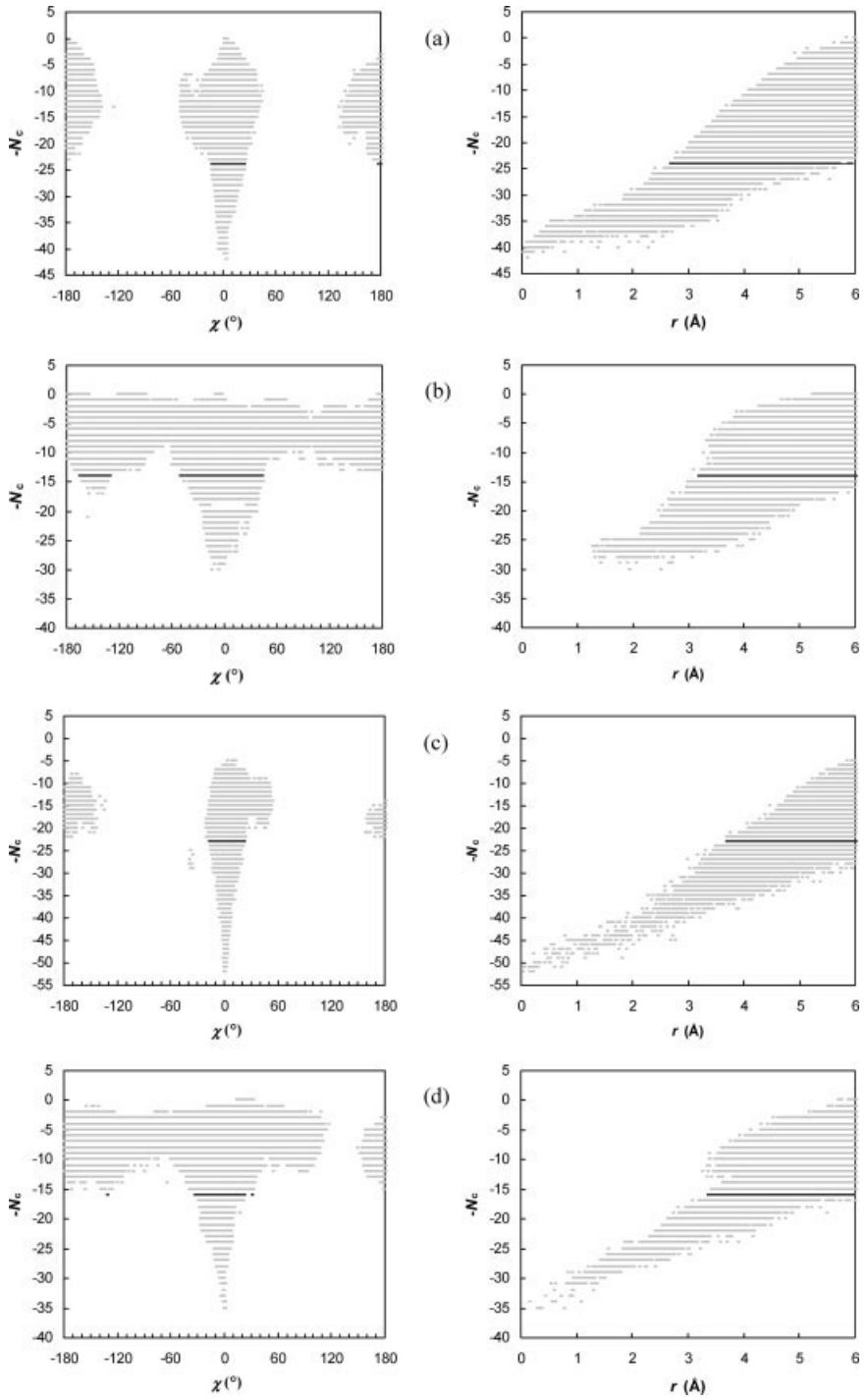

Figure 4

*Scatter plots of the contact level N_c versus the rotation angle χ or the relative separation r. For clarity, the full range of χ, from $-180°$ to $180°$, is divided into 500 bins and, at each contact level, at most one sampled χ value is saved for displaying. The sampled configurations are selected in an analogous manner for displaying in the N_c-r scatter plot, with the sampled range of r, from 0 to 6 Å, divided into 500 bins. The N_c-χ and N_c-r pairs of plots labeled (**a**)–(**d**) are for the E9:Im9, Bn:Bs, AChE:Fas, and IL4:IL4BP complexes, respectively.*

with the dielectric boundary specified as the vdW surface. In Figure 5, we show that the agreement between theory and experiment for the salt dependences of the association rates is further improved when $\langle U_{el} \rangle^*$ is calculated by the nonlinear PB equation, again with the dielectric boundary specified as the vdW surface. The

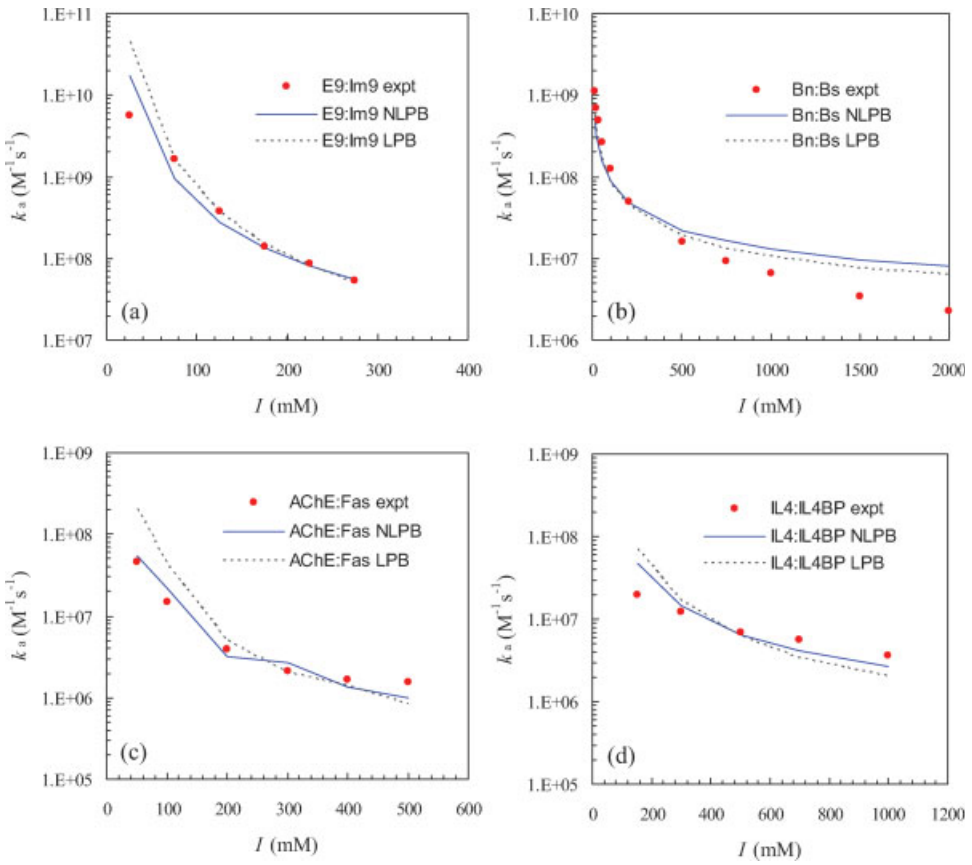

Figure 5

*Salt dependences of the association rates of the four protein complexes listed on (**a**)–(**d**). Experimental results are shown in circles; predictions by the nonlinear (NLPB) and linearized (LPB) Poisson–Boltzmann equations are shown in solid and dotted curves, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

overestimations at low ionic strengths for the E9:Im9, AChE:Fas, and IL4:IL4BP complexes by the linearized PB results are now reduced by the nonlinear PB results. For the Bn:Bs complex, the difference between linearized and nonlinear PB results is small.

As reported previously,[21] in the comparison between experimental data and the linearized PB results for the salt dependences of the association rates, the basal rate k_{D0} is treated as a fitting parameter. The fitted values are 5×10^5, 1.4×10^6, 5×10^4, and 5×10^4 M^{-1} s^{-1}, respectively, for E9:Im9, Bn:Bs, AChE:Fas, and IL4-IL4BP complexes. With $\langle U_{el} \rangle^*$ calculated by the nonlinear PB equation, small upward adjustments in the basal rate are required, leading to values at 9×10^5, 2×10^6, 9×10^4, and 10^5 M^{-1} s^{-1} for the four complexes. It is important to note that these values fall in the range of 10^5–10^6 M^{-1} s^{-1} expected of the basal rate.[1–3] While these k_{D0} values are consistently higher than those obtained from Brownian dynamics simulations (as already noted and rationalized), it is interesting that among both sets the Bn:Bs complex has the highest k_{D0} values.

Mutation effects on association rates

In Figure 6, we show that the improvement in agreement between theory and experiment upon calculating $\langle U_{el} \rangle^*$ by the nonlinear PB equation also extends to the 23 mutants of the 4 protein complexes. The fitted values of k_{D0} reported earlier are used on the mutants without further adjustment. For most of these mutants, studied over various ionic strengths, the nonlinear PB results move from the linear PB results modestly toward the corresponding experimental results. This can be seen, e.g., on the E9Q mutant of the IL4:IL4BP complex, the E60A mutant of the Bn:Bs complex, and the D283N mutant of the AChE:Fas complex.

Electrostatic rate retardation predicted by PB-MS calculations

In previous studies on electrostatic contributions to the binding energies of proteins, it was found that, when PB calculations with the vdW surface give negative values for the electrostatic interaction energies of oppositely

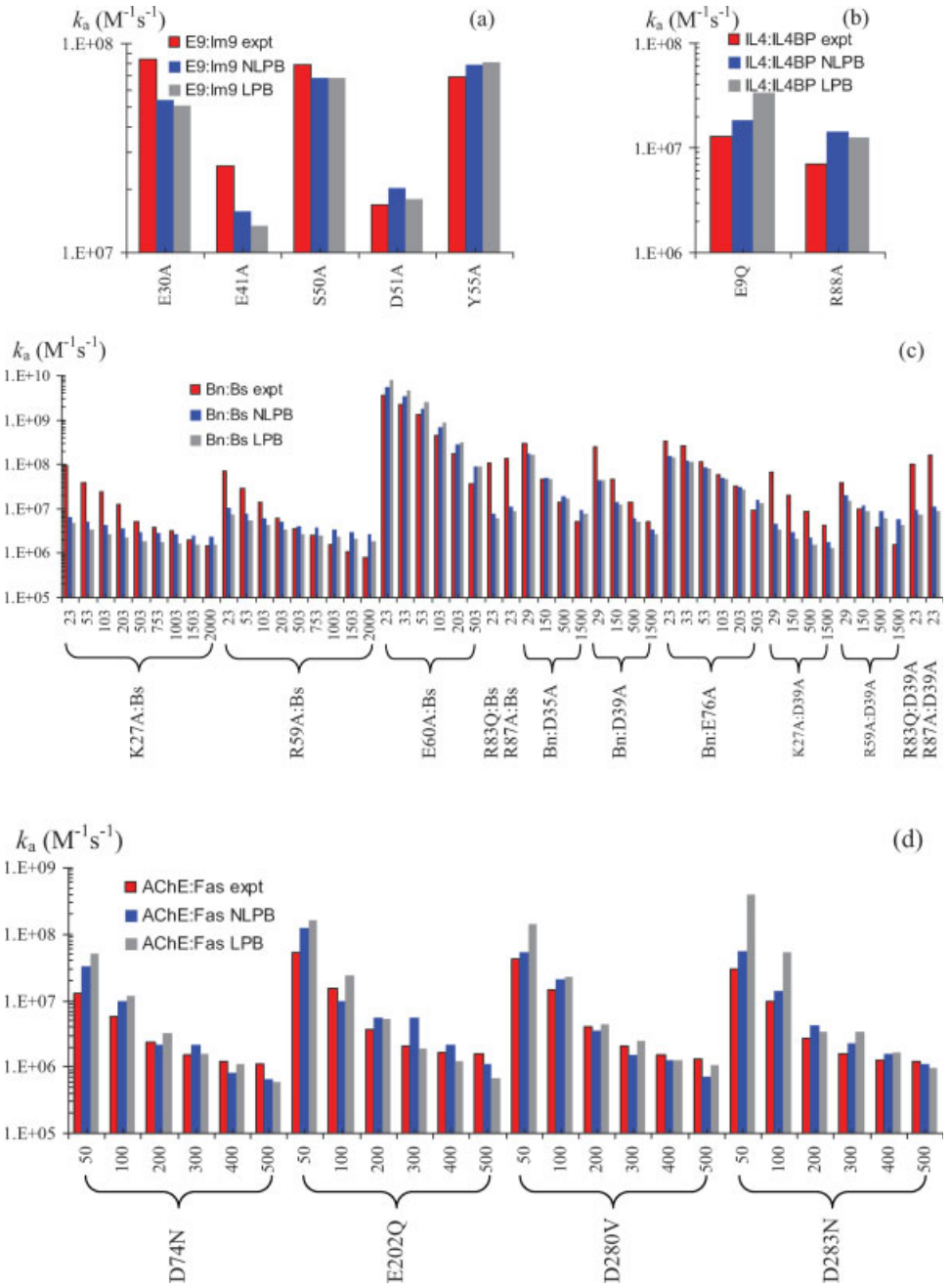

Figure 6

*Mutation effects on the association rates of the four protein complexes. Experimental and predicted (NLPB and LPB) rates are shown as red, blue, and black bars, respectively. (**a**) Rates for five Im9 mutations in the E9:Im9 complex at I = 225 mM. (**b**) Rates for two IL4 mutations in the IL4:IL4BP complex at I = 150 mM. (**c**) Rates for 12 mutations on the Bn:Bs complex at the various ionic strengths (in mM) indicated. (**d**) Rates for four AChE mutations in the AChE:Fas complex at the various ionic strengths (in mM) indicated. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

charged subunits, the electrostatic interaction energies can become positive when the MS is used.[50,51,61,62] Here we find the same sign reversal on the transient-complex electrostatic interaction energies of all the four protein complexes. When the dielectric boundary is changed from the vdW surface to MS, the sign of $\langle U_{el} \rangle^*$ changes from negative to positive, leading to electrostatic rate retardation instead of enhancement. For example, for the Bn:Bs complex, when the ionic strength is varied from 13 to 2000 m*M*, $\langle U_{el} \rangle^*$ calculated by the nonlinear

PB equation with the vdW surface varies from −3.30 to −0.82 kcal mol^{-1}. Correspondingly, $\langle U_{el}\rangle^*$ calculated by the nonlinear PB equation with MS varies from 2.50 to 5.13 kcal mol^{-1}. For the latter results to be consistent with the experimental data for the association rate would require a basal rate in the order of 10^{10}–10^{11} M^{-1} s^{-1}, which clearly seems unphysical.

DISCUSSION

We have demonstrated the predictive power of the transient-complex theory for association rates on four protein complexes and a large number of their mutants. We found that use of the nonlinear PB equation instead of the less rigorous, linearized version in calculating electrostatic interaction energies leads to improved agreement with experiment. We also found that specifying the dielectric boundary as the MS in the electrostatic calculations leads to rate retardation instead of enhancement, an apparently unphysical result.

Overview on the spectrum of protein–protein association rates

As illustrated in Figure 1, protein–protein association rates span many orders of magnitude. Rates toward the high end are limited by diffusion; rates toward the low end are limited by conformational changes that accompany complex formation. The demarcation point between these two regimes lies probably around 10^5 M^{-1} s^{-1}. In the diffusion-limited regime, association rates for proteins that do not involve significant electrostatic contributions, such as antibodies and protein antigens, fall within the narrow range of 10^5–10^6 M^{-1} s^{-1}. Higher association rates implicate electrostatic enhancement, resulting from complementary electrostatic surfaces.

A value in the range of 10^9–10^{10} M^{-1} s^{-1} calculated from the Smoluchowski result $4\pi DR$ has sometimes been used as a benchmark for determining whether protein–protein association is diffusion-limited. As we have emphasized, the Smoluchowski model is inappropriate for stereospecific association. With more realistic models such as spheres with stereospecific contacts, one arrives at a basal rate for diffusion-limited association in the range of 10^5–10^6 M^{-1} s^{-1}. This range should be used as the basis for whether to propose diffusion control for the association of a particular pair of proteins.

In general, rates beyond the 10^5–10^6 M^{-1} s^{-1} range implicate electrostatic enhancement. A hallmark of electrostatically enhanced diffusion-limited association is manifested by disparate ionic-strength effects on the association rate k_a and the dissociation rate k_d.[19] Specifically, k_a decreases significantly with increasing ionic strength, but k_d is affected by ionic strength only marginally. A sample of 10 complexes, including the four studied here, that exhibit such a telltale sign for electrostatically enhanced diffusion-limited association are listed in a recent paper.[63] The disparate ionic-strength effects on k_a and k_d provided strong experimental support for the present transient-complex theory, because these effects are precisely what the theory predicts.[19] Since the transient complex lies at the outer boundary of the bound state and is very closer to the latter, ionic strength will affect the electrostatic interaction energies of the transient complex and the bound state to similar extents. Hence ionic strength will affect k_a almost as strongly as the association constant K_a. The dissociate rate k_d, given by k_a/K_a, will therefore only be weakly affected by ionic strength. In the present paper, we have gone beyond qualitative explanation and shown that ionic-strength effects on k_a can be quantitatively predicted by the transient-complex theory.

Absence of energy barrier for diffusion-limited protein–protein association

Of the host of theoretical models for diffusion-limited association discussed in this paper, none requires an energy barrier. In the Smoluchowski model, the contact surface $r = R$ is absorbing; beyond it the energy landscape is flat. The Debye model introduces an energy function such as the Coulomb type, $-Q/r$, that for practical purposes most likely will give rise to a funnel-like landscape. The models for stereospecific association involve a dividing surface that is reflecting everywhere, except for a small absorbing patch. Outside the dividing surface, the energy landscape is either flat, as in the case of force-free diffusion (leading to the basal rate k_{D0}), or funnel-like when electrostatic rate enhancement is involved. Any local minimum in the energy surface has no particular significance in determining the diffusion-limited association rate, k_D. Inside the dividing surface, i.e., within the bound state, there may well be energy barriers, but these have no bearing on k_D.

When a flat or funnel-like multidimensional energy surface is reduced to a potential of mean force in a single "reaction coordinate," an entropy barrier may arise.[64,65] In treating diffusion-limited stereospecific association, involving six translational and rotational degrees of freedom, it does not seem to make much sense to identify a single reaction coordinate. Reduction to two coordinates, i.e., r and χ, appears to have some merit, since these are the only coordinates that exhibit sharp transitions in the ranges of sampled values on going from the bound state to the unbound state. In our previous study, we found that the potential of mean force in r and χ is still funnel-like, without a barrier.[20] It may be tempting to regard the orientational constraints of a stereospecific association model, leading to a four order of magnitude reduction in basal rate relative to the Smoluchowski model, as an entropy barrier. However, that does not provide any

additional insight into the diffusional process leading to the bound state, nor does it lead to a quantitative formulation of the association rate. Our transient-complex theory, in the form of Eq. (4), accomplishes that.

Modulation of protein association rate by electrostatic interactions

The transient-complex electrostatic interaction energy, $\langle U_{el} \rangle^*$, has values as negative as -5.8 kcal mol^{-1} among the four protein complexes studied. That, according to Eq. (4), corresponds to a rate enhancement of 10^4-fold. It should be noted that such a dramatic rate enhancement is only possible for stereospecific association. In the Debye model, for example, for a Coulomb potential $-Q/r$, with the contact value (i.e., $-Q/R$) at -5.8 kcal mol^{-1}, gives merely a 10-fold rate enhancement.

The remarkable ability of electrostatic interactions in modulating protein association rates is demonstrated by the 10^3-fold difference between two Rho GTPases, Cdc42 and TC10, in the rates of associating with the Wiskott-Aldrich syndrome protein.[11] Cdc42 and TC10 share 70% sequence identity. The large difference in association rates between related proteins suggests that association rate, in addition to binding affinity, can provide a mechanism for specificity. Other possible biological roles of association rates have also been proposed.[21]

The ability of electrostatic interactions to contribute up to 10^4-fold rate enhancement is directly related to their long-range nature, which, it should be recalled, is one of the two conditions for the validity of Eq. (4). As shown by calculations on model systems, short-range interactions have much less effects on the association rate.[66] In a previous study we found that short-range interactions are present in the transient complex.[20] More specifically, the transient complex is surrounded by a broad shallow basin, arising from one or a few loosely formed native interactions. Such interactions may contribute to a small enhancement in the association rate. If this effect from the short-range interactions is taken into consideration, the basal rate required to fit with experimental data will be reduced, bringing it into closer agreement with the value of k_{D0} obtained from Brownian dynamics simulations.

Calculation of electrostatic rate enhancement

Our assessment of the electrostatic contributions to association rates obtained from four types of PB calculations against experimental data demonstrates that details of treating electrostatic interactions are important. This is especially true when the two proteins are close. In studies of electrostatic rate enhancement by Brownian dynamics simulations, the prohibitive cost of calculating forces and torques from a rigorous implementation of the PB equation on the fly led to approximate treatments, such as those using test charges or effective charges.[29–32] The errors of such treatments can be significant when two proteins are in close proximity, leading to uncertainty about predicted electrostatic rate enhancement. Our transient-complex theory appears to provide the only viable option for rigorous treatment of electrostatic interactions in studying protein association.

As noted already, the calculated electrostatic interaction energy is sensitive to the precise specification of the dielectric boundary.[50,51,61,62,67,68] Past efforts in discriminating between two choices, vdW and MS, have relied on experimental data for mutational effects on folding stability or binding affinity. In this paper we make use of experimental data on association rates, which offer two important advantages. First, as presented by earlier arguments, rate enhancement is dominated by long-ranged electrostatic interactions; thus experimental data on association rates are not appreciably "contaminated" by short-range nonelectrostatic effects. In the past, comparison of calculated electrostatic contributions to experimental data on binding affinity, which may contain nonelectrostatic contributions, always raised concerns, which bring out the second advantage of using experimental data on association rates. That is, in the past, in order to cancel nonelectrostatic contributions in experimental data on binding affinity, comparisons have been restricted to relative effects of mutations. The problem is that the relative effects of mutations obtained by vdW and MS calculations are often very similar. Here we test calculated results against experimental data for absolute rates instead of relative rates. Our results show that vdW calculations lead to electrostatic rate enhancement, but MS calculations lead to rate retardation. We suggest that the latter is unphysical.

REFERENCES

1. Northrup SH, Erickson HP. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. Proc Natl Acad Sci USA 1992;89:3338–3342.
2. Zhou H-X. Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. Biophys J 1997;73:2441–2445.
3. Schlosshauer M, Baker D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Sci 2004; 13:1660–1669.
4. Foote J, Eisen HN. Kinetic and affinity limits on antibodies produced during immune responses. Proc Natl Acad Sci USA 1995;92: 1254–1256.
5. Hoffman TL, LaBranche CC, Zhang W, Canziani G, Robinson J, Chaiken I, Hoxie JA, Doms RW. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. Proc Natl Acad Sci USA 1999;96:6359–6364.
6. Wassaf D, Kuang G, Kopacz K, Wu QL, Nguyen Q, Toews M, Cosic J, Jacques J, Wiltshire S, Lambert J, Pazmany CC, Hogan S, Ladner RC, Nixon AE, Sexton DJ. High-throughput affinity ranking of antibodies using surface plasmon resonance microarrays. Anal Biochem 2006;351:241–253.

7. Wallis R, Moore GK, James R, Kleanthous C. Protein-protein interactions in colicin E9 DNase-immunity protein complexes. I. Diffusion-controlled association and femtomolar binding for the cognate complex. Biochemistry 1995;34:13743–13750.
8. Schreiber G, Fersht AR. Rapid, electrostatically assisted association of proteins. Nat Struct Biol 1996;3:427–431.
9. Terlau H, Shon K-J, Grilley M, Stocker M, Stuhmer W, Baldomero OM. Strategy for rapid immobilization of prey by a fish-hunting marine snail. Nature 1996;381:148–151.
10. Radic Z, Kirchhoff PD, Quinn DM, McCammon JA, Taylor P. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. J Biol Chem 1997;272:23265–23277.
11. Hemsath L, Dvorsky R, Fiegen D, Carlier MF, Ahmadian MR. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. Mol Cell 2005;20:313–324.
12. Marchand J-B, Kaiser DA, Pollard TD, Higgs HN. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. Nat Cell Biol 2001;3:76–82.
13. Shen BJ, Hage T, Sebald W. Global and local determinants for the kinetics of interleukin-4/interleukin-4 receptor α chain interaction. A biosensor study employing recombinant interleukin-4-binding protein. Eur J Biochem 1996;240:252–261.
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–10041.
15. Zhou H-X. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. Biophys J 1993;64:1711–1726.
16. Zhou H-X. Effect of interaction potentials in diffusion-influenced reactions with small reactive regions. J Chem Phys 1996;105:7235–7237.
17. Zhou HX, Wong KY, Vijayakumar M. Design of fast enzymes by optimizing interaction potential in active site. Proc Natl Acad Sci USA 1997;94:12372–12377.
18. Vijayakumar M, Wong K-Y, Schreiber G, Fersht AR, Szabo A, Zhou H-X. Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. J Mol Biol 1998;278:1015–1024.
19. Zhou H-X. Disparate ionic-strength dependencies of on and off rates in protein-protein association. Biopolymers 2001;59:427–433.
20. Alsallaq R, Zhou H-X. Energy landscape and transition state of protein-protein association. Biophys J 2007;92:1486–1502.
21. Alsallaq R, Zhou H-X. Prediction of protein-protein association rates from a transition-state theory. Structure 2007;15:215–224.
22. Smoluchowski MV. Versuch einer mathematischen theorie der koagulationskinetik kolloider losungen. Z Phys Chem 1917;92:129–168.
23. Debye P. Reaction rates in ionic solutions. Trans Electrochem Soc 1942;82:265–272.
24. Solc K, Stockmayer WH. Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. II. Approximate steady-state solution. Int J Chem Kinet 1973;5:733–752.
25. Berg OG. Orientation constraints in diffusion-limited macromolecular association. The role of surface diffusion as a rate enhancing mechanism. Biophys J 1985;47:1–14.
26. Shoup D, Lipari G, Szabo A. Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. Biophys J 1981;36:697–714.
27. Temkin SI, Yakobson BI. Diffusion-controlled reactions of chemically anisotropic molecules. J Phys Chem 1984;88:2679–2682.
28. Schlosshauer M, Baker D. A general expression for bimolecular association rates with orientational constraints. J Phys Chem B 2002; 106:12079–12083.
29. Northrup SH, Reynolds JCL, Miller CM, Forrest KJ, Boles JO. Diffusion-controlled association rate of cytochrome *c* and cytochrome *c* peroxidase in a simple electrostatic model. J Am Chem Soc 1986;108:8162–8170.
30. Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. Biophys J 1997;72:1917–1929.
31. Elcock AH, Gabdoulline RR, Wade RC, McCammon JA. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. J Mol Biol 1999;291:149–162.
32. Gabdoulline RR, Wade RC. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. J Mol Biol 2001;306:1139–1155.
33. De Rienzo F, Gabdoulline RR, Menziani MC, De Benedetti PG, Wade RC. Electrostatic analysis and Brownian dynamics simulation of the association of plastocyanin and cytochrome *f*. Biophys J 2001;81:3090–3104.
34. Zou G, Skeel RD. Robust biased Brownian dynamics for rate constant calculation. Biophys J 2003;85:2147–2157.
35. Lin J, Beratan DN. Simulation of electron transfer between cytochrome C2 and the bacterial photosynthetic reaction center: Brownian dynamics analysis of the native proteins and double mutants. J Phys Chem B 2005;109:7529–7534.
36. Haddadian EJ, Gross EL. A Brownian dynamics study of the interactions of the luminal domains of the cytochrome b6f complex with plastocyanin and cytochrome c6: the effects of the Rieske FeS protein on the interactions. Biophys J 2006;91:2589–2600.
37. Spaar A, Dammer C, Gabdoulline RR, Wade RC, Helms V. Diffusional encounter of barnase and barstar. Biophys J 2006;90:1913–1924.
38. Zhou H-X, Briggs JM, McCammon JA. A 240-fold electrostatic rate-enhancement for acetylcholinesterase-substrate binding can be predicted by the potential within the active site. J Am Chem Soc 1996;118:13069–13070.
39. Zhou H-X, Briggs JM, Tara S, McCammon JA. Correlation between rate of enzyme-substrate diffusional encounter and average Boltzmann factor around active site. Biopolymers 1998;45:355–360.
40. Shoup D, Szabo A. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. Biophys J 1982;40:33–39.
41. Eyring H. The activated complex in chemical reactions. J Chem Phys 1932;3:107–115.
42. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J Mol Biol 2000;301:1163–1178.
43. Buckle AM, Schreiber G, Fersht AR. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. Biochemistry 1994;33:8878–8889.
44. Bourne Y, Taylor P, Marchot P. Acetylcholinesterase inhibition by fasciculin: crystal structure of the complex. Cell 1995;83:503–512.
45. Hage T, Sebald W, Reinemer P. Crystal structure of the interleukin-4/receptor α chain complex reveals a mosaic binding interface. Cell 1999;97:271–281.
46. Schreiber G, Fersht AR. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry 1993;32:5145–5150.
47. Schreiber G, Fersht AR. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. J Mol Biol 1995;248:478–486.
48. Frisch C, Fersht AR, Schreiber G. Experimental assignment of the structure of the transition state for the association of barnase and barstar. J Mol Biol 2001;308:69–77.
49. Wallis R, Leung K-Y, Osborne MJ, James R, Moore GR, Kleanthous C. Specificity in protein-protein recognition: conserved Im9 residues are the major determinations of stability in the colicin E9 Dnase-Im9 complex. Biochemistry 1998;37:476–485.
50. Dong F, Vijayakumar M, Zhou H-X. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. Biophys J 2003;85:49–60.
51. Dong F, Zhou H-X. Electrostatic contribution to the binding stability of protein-protein complexes. Proteins 2006;65:87–102.

52. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput Phys Commun 1995;91:57–95.
53. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
54. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118:11225–11236.
55. Bondi A. van der Waals volumes and radii. J Phys Chem 1964; 68:441–451.
56. Zhou H-X, Szabo A. Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions. Biophys J 1996; 71:2440–2457.
57. Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. J Chem Phys 1978;69:1352–1360.
58. Fernandes MX, de la Torre JG. Brownian dynamics simulation of rigid particles of arbitrary shape in external fields. Biophys J 2002;83:3039–3048.
59. Zhou H-X. Calculation of translational friction and intrinsic viscosity. II. Application to globular proteins. Biophys J 1995;69:2298–2303.
60. Zhou HX. Kinetics of diffusion-influenced reactions studied by Brownian dynamics. J Phys Chem 1990;94:8794–8800.
61. Wang T, Tomic S, Gabdoulline RR, Wade RC. How optimal are the binding energetics of barnase and barstar? Biophys J 2004;87:1618–1630.
62. Qin SB, Zhou HX. Do electrostatic interactions destabilize protein-nucleic acid binding? Biopolymers 2007;86:112–118.
63. Zhou H-X. Association and dissociation kinetics of colicin E3 and immunity protein 3: convergence of theory and experiment. Protein Sci 2003;12:2379–2382.
64. Zhou H-X, Zwanzig R. A rate process with an entropy barrier. J Chem Phys 1991;94:6147–6152.
65. Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. Protein Sci 2000;9:452–465.
66. Zhou HX, Szabo A. Enhancement of association rates by nonspecific binding to DNA and cell membranes. Phys Rev Lett 2004; 93:178101.
67. Vijayakumar M, Zhou HX. Salt bridges stabilize the folded structure of barnase. J Phys Chem B 2001;105:7334–7340.
68. Dong F, Zhou H-X. Electrostatic contributions to T4 lysozyme stability: solvent-exposed charges versus semi-buried salt bridges. Biophys J 2002;83:1341–1347.