CHEMICAL REVIEWS

Volume 109, Number 3

Fundamental Aspects of Protein−Protein Association Kinetics

G. Schreiber,*,† G. Haran,*,‡ and H.-X. Zhou*,§

Departments of Biological Chemistry and Chemical Physics, Weizmann Institute of Science, Rehovot, 76100, Israel and Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

Received May 22, 2008

Contents

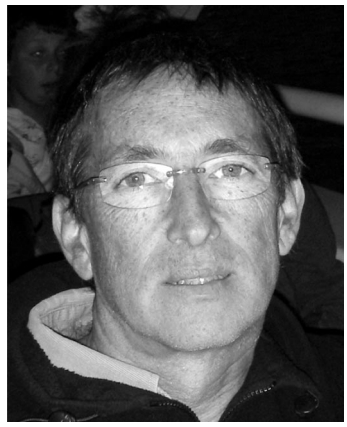
* To whom correspondence should be addressed. G.S.: phone, +972-8-9343249; fax, +972-8-9346095; e-mail, gideon.schreiber@weizmann.ac.il. G.H.: phone, +972-8-9342625; fax, +972-8-9342749; e-mail, gilad.haran@weizmann.ac.il. H.-X.Z.: phone, 850-6451336; fax, 850-6447244; e-mail, hzhou4@fsu.edu.
† Department of Biological Chemistry, Weizmann Institute of Science.
‡ Department of Chemical Physics, Weizmann Institute of Science.
§ Florida State University.

1. Introduction

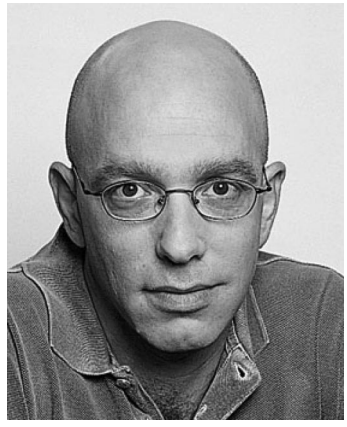
The structure of a protein complex together with information about its affinity and other thermodynamic characteristics provide a "frozen" view of the complex. This picture ignores the kinetic nature of protein−protein association and dissociation, which is of major biological and biophysical interest. This review focuses on recent advances in deciphering the kinetic pathway of protein complex formation, the nature of the precomplex formed through diffusion (which we termed the "transient complex"[1]), the transition state, and other intermediates (such as the so-called encounter complex) along the association pathway.

Protein−protein association is at the center of diverse biological processes ranging from enzyme catalysis/inhibition to regulation of immune response by cytokines. The association rates often play a critical role in such processes, as in situations where speed is of essence.[2] For example, the purple cone snail and other venomous animals capture prey with remarkable efficiency and speed by releasing toxins that rapidly bind to ion channels;[3] the green mamba achieves a similar feat by targeting acetylcholinesterase (AChE), an enzyme essential for the integrity of neural transmission.[4] Bacteria such as *Escherichia coli* and *Bacillus amyloliquefaciens* excrete nucleases as weapons against competitors or predators. Defense of the producing cells from damage to their own DNA or RNA by such nucleases requires rapid association with cognate inhibitors.[5,6] Indeed, in the last example rapid association is such a priority that the inhibitor barstar has a cluster of acidic residues that facilitate association with the nuclease barnase, even though the clustered charges reduce folding stability.[7] In the ruminant gut RNase

10.1021/cr800373w CCC: $71.50
Published on Web 02/05/2009



Gideon Schreiber received his Doctorate in Biochemistry at the Hebrew University in Jerusalem in 1992. After a postdoctoral period at the MRC in Cambridge, U.K., at the laboratory of Professor Alan Fersht, he joined the Weizmann Institute of Science as Senior Scientist. Presently, he is an associate professor at the Department of Biological Chemistry at the same institute. His research interest is in the investigation of protein−protein interactions from the basic understanding of the mechanism to protein design. His work spans from bioinformatics and algorithm development, biophysical bench work, protein design and engineering, to applied biology. In addition he is a director of the Israel Structural Proteomic Center, whose aim is to provide structures of proteins and protein complexes to the biological community.



Gilad Haran is currently an associate professor in the Department of Chemical Physics of the Weizmann Institute of Science. He did his Ph.D. work at the Weizmann Institute with Professors Ephraim Katchalsky-Katzir and Elisha Haas. He was then a postdoctoral fellow with Professor Robin Hochstrasser at the University of Pennsylvania. His laboratory is using single-molecule spectroscopy to study molecular dynamics, from protein folding and association to charge-transfer dynamics on metal surfaces.

A is required for degrading accumulated RNA; potential toxicity of leaked nuclease is prevented by rapid association with a ribonuclease inhibitor.[8,9]

Reorganization of the Actin cytoskeleton provides yet another illustration of the importance of rapid protein association. Reorganization is attained through Actin polymerization, which is nucleated by the Arp2/3 complex. The latter is activated by the Wiskott−Aldrich Syndrome protein (WASp), which in turn is released from the autoinhibited state by the Rho GTPase Cdc42.[10] As Actin polymerization is initiated with a nucleation process, the speed of upstream signaling has a critical impact on the rate of polymer formation. It is thus not surprising that high association rate constants have been observed between partners along the signaling pathway.[11,12] The high association rate constant between Cdc42 and WASp has been found to be essential for the latter to stimulate Actin polymerization as another

Huan-Xiang Zhou received his Ph.D. degree from Drexel University in 1988. He did postdoctoral work at the NIH with Attila Szabo. After faculty appointments at HKUST and Drexel he moved to Florida State University in 2002. His group does theoretical, computational, and experimental research on protein−protein interactions, crowding and confinement effects of cellular environments, and biological roles of protein dynamics. He is currently the Editor-in-Chief of *PMC Biophysics*.

Rho GTPase sharing 70% sequence identity, TC10, with an identical dissociation rate constant but a 1000-fold lower association rate constant failed to stimulate Actin polymerization.[11] The failure to stimulate Actin polymerization in patients carrying mutant WAS genes is the root cause of the Wiskott−Aldrich Syndrome.

Several other compelling arguments can be made for the biological roles of rapid protein association.[13] (a) Fast association may enhance binding affinity. High affinity can also be achieved through slow dissociation; however, for proteins involved in signaling slow dissociation is not an option since it implies a long-lasting bound state, which effectively corresponds to a permanent off or on switch. A good example for this is the binding of Ras to its natural affector Raf. This protein dissociates within a fraction of a second but maintains an affinity in the nanomolar range through fast association. Moreover, the difference between the natural effector, Raf, and the non-natural effector, Ral, lies in their rates of association with Ras.[14] Therefore, even if not for a direct reason (such as in stimulation of Actin polymerization), the affinity requirement alone may call for fast association. (b) Enzyme−substrate binding is a determining factor for the overall turnover rate and becomes *the* rate-limiting step for catalytically "perfect" enzymes. Substrate-binding rate constants of such enzymes reach 10^8 M^{-1} s^{-1} and beyond, as found for the ribotoxin restrictocin and RNase A.[15,16] (c) When several proteins compete for the same receptor or when one protein is faced with alternative pathways, kinetic control, not thermodynamic control, dominates in many cases; this is especially true when dissociation is slow. For example, during protein synthesis cognate and noncognate aminoacyl-tRNA synthetases can potentially compete for the same tRNA. As an additional example, consider newly synthesized proteins, which potentially face aggregation if not isolated by a chaperone. From the point of view of kinetic control, it is easy to see why rapid binding of denatured proteins to the chaperonin GroEL has been observed.[17] (d) Differences in binding rate between related proteins may serve as an additional mechanism for specificity, as can be suggested for Rho GTPases Cdc42 and TC10 and for Ras effectors Raf and Ral.

The examples and arguments presented above suggest that rapid binding is as important as high affinity in the proper

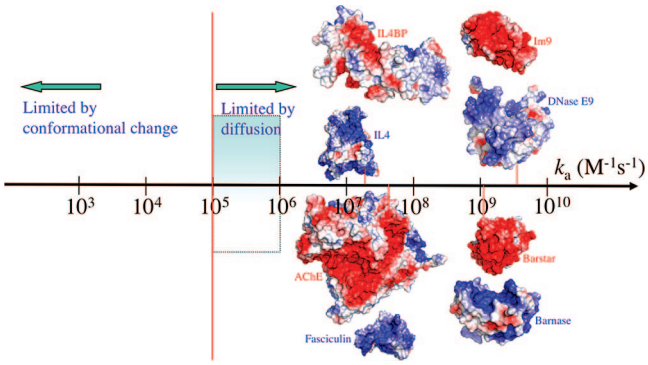


Figure 1. Wide spectrum of association rate constants. The red vertical line marks the start of the diffusion-controlled regime. The shaded range marks the absence of long-range forces. (Adapted with permission from ref 1. Copyright 2008 Wiley Interscience.)

functioning of proteins. It is now increasingly recognized that proteins function in the context of multicomponent complexes. Manipulating association rate constants of various components presents unique opportunities for the control of protein functions. Many interactions between proteins are also targeted for drug development; in designing such drugs, both high affinity and rapid binding should be taken into consideration.

1.1. Overview of Protein Association Kinetics

The observed rate constants of protein association span a wide range from $<10^3$ to $>10^9$ M^{-1} s^{-1} (Figure 1). In comprehending these values a basic fact is that for two proteins to recognize each other their interfaces have to be oriented with high specificity. A relative rotation of as little as a few degrees or a relative translation by a few Angstroms is sufficient to break all specific interactions between the two proteins.[18] The rate of association of a protein complex is limited by diffusion and geometric constraints of the binding sites and may be further reduced by subsequent chemical processes.[19]

To better understand the kinetics of association of two proteins (A and B) it is useful to consider the process as going through an intermediate state (A*B) in which the two proteins have near-native separations and orientations.[1,20−23] We refer to this intermediate state as the transient complex,[1,20] noting that it is sometimes also termed the encounter complex.[24] A more detailed discussion of terminology as well as the specification of the ensemble of configurations making up the transient complex is provided in section 3. From this ensemble, conformational rearrangement can lead to the native complex (C). Accordingly we have the kinetic scheme

$$A + B \underset{k_{-D}}{\overset{k_D}{\rightleftharpoons}} A^*B \xrightarrow{k_c} C \qquad (1)$$

While the first step of this scheme depends on relative diffusion between the protein molecules, the second step is akin to an intramolecular chemical reaction and can therefore be described by the classical transition-state theory[25] (with the transition state located at the top of the free-energy barrier separating A*B from C[26]) or by Kramers' theory.[27] The latter theory accounts for barrier recrossing and models motion along the reaction coordinate as diffusive. The overall rate constant of association is

$$k_a = \frac{k_D k_c}{k_{-D} + k_c} \qquad (2)$$

which is bounded by the diffusion-controlled rate constant, k_D, for reaching the transient complex. This limit is reached when conformational rearrangement is fast relative to the dissociation of the transient complex (i.e., $k_c \gg k_{-D}$), leading to

$$k_a \approx k_D \qquad (3a)$$

In the opposite limit $k_c \ll k_{-D}$, conformational rearrangement or reaction becomes rate limiting and

$$k_a \approx k_c k_D / k_{-D} \equiv k_R \qquad (3b)$$

Note that k_D/k_{-D} is the equilibrium constant for forming the transient complex.

There is no simple test that would place the association of a given protein complex into the diffusion-controlled, reaction-controlled, or mixed regime. However, two general statements can be made. First, a diffusion-controlled rate constant falls on the high end of the spectrum of observed values and a reaction-controlled rate constant falls on the low end (Figure 1). Second, diffusion-controlled association typically involves only local conformational changes between the unbound proteins and the native complex whereas reaction-controlled association typically involves gross changes such as loop reorganization or domain movement. These points will be further amplified below. For later reference we also introduce here the overall dissociation rate constant, k_d, and the overall association constant, $K_a = k_a/k_d$.

In the transient complex the two protein molecules must satisfy translational/rotational constraints, which severely hinder the diffusion-controlled rate constant k_D. In the absence of any biasing force, theoretical estimates put the resulting "basal" value, k_{D0}, in the range from 10^5 to 10^6 M^{-1} s^{-1} (see below).[28−31] Antibody−protein association rate constants are typically observed in this narrow range.[32−34] The value 10^5 M^{-1} s^{-1} thus may mark the start of the diffusion-controlled regime.

To go beyond the basal rate constant k_{D0} and reach values in the range from 10^8 to 10^9 M^{-1} s^{-1} as observed for many protein complexes[3−6,8,9,15−17,35,36] (Figure 1), intermolecular forces must be present. For a force to speed up a "diffusion-controlled association" it must be present in the diffusion process that leads to the transient complex. Indeed, analytical results on model systems show that when the range of the force is reduced the resulting rate enhancement decreases drastically.[29,37,38] For protein−protein association the dominant long-range force is provided by electrostatic interactions.

Rate constants higher than k_{D0} are indeed usually correlated with favorable electrostatic interactions, as manifested by complementary charge distributions on the binding partners, which are illustrated in Figure 1 for four protein pairs. RNA and DNA have negative charges from phosphate groups; proteins targeting them (like barnase, RNase A, and DNase E9) generally have enriched distributions of basic residues on the nucleic acid binding sites.[39,40] Electrostatically enhanced protein−nucleic acid association rates can thus be anticipated. The basic residues on the nucleic acid binding sites can also be easily exploited by inhibitors (like barstar, ribonuclease inhibitor, and Im9): fast inhibition can be achieved through a concentration of acidic residues on the latter molecules. Toxins blocking the *Shaker* potassium

channel follow a similar strategy.[23] Apparently for facilitating the conduction of the positively charged potassium ion the mouth of the channel pore is lined with two rings of acidic residues.[41] To complement the resulting negative electrostatic surface channel toxins have excess basic residues. The charge-complementarity argument can also be made to rationalize the excess basic residues on fasciculin 2 (fas), the snake toxin targeting AChE. AChE uses a negative electrostatic surface around the entrance to the active-site gorge for the fast binding of its positively charged substrate.[4,42] In other cases, such as in complexes formed by WASp and Cdc42[43] and by interleukin-4 (IL4) and its receptor,[44] the reasons for a particular subunit to take up either positive or negative charges are not obvious. That they nonetheless show charge complementarity implicates functional roles of their fast association.

1.2. Analytical Models of Protein Association Rate Constants

The first theoretical result for the association rate constant was obtained by Smoluchowski,[45] who found that the diffusion-controlled rate constant for two uniformly reacting spheres to form a complex is

$$k_{D0} = 4\pi DR \tag{4}$$

where D is the relative translational diffusion constant and R is the contact distance between the centers of the two spheres. Debye recognized that the association rate between oppositely charged molecules can be increased by electrostatic interactions.[46] For two uniformly reacting spheres with a centrosymmetric interaction potential $U(r)$ he found the diffusion-controlled rate constant to be

$$k_D = 4\pi D / \int_R^\infty e^{U(r)/k_BT} r^{-2}\, dr \tag{5}$$

where r is the intersphere distance and k_BT is thermal energy. In the following discussion we will use k_{D0} to denote the diffusion-controlled rate constant in the absence of an interaction potential (also referred to as the basal rate constant) and use k_D to denote the counterpart in the presence of an interaction potential. The enhancement over the basal rate constant by an attractive interaction potential in the Debye model is quite modest. For example, for a Coulombic interaction potential $U(r) = -Q/r$ one finds, upon evaluating the integral in eq 5, $k_D/4\pi DR = (Q/k_BTR)/(1 - e^{-Q/kBTR})$. The rate enhancement, $k_D/4\pi DR$, is 9-fold when the magnitude of the potential is $9k_BT$ at contact. This modest rate enhancement will be contrasted below with a much greater enhancement predicted on a more realistic model for protein association.

For medium-sized proteins, R and D are on the order of 40 Å and 20 Å^2 ns^{-1}, respectively. Equation 4 would predict a basal rate constant of 6×10^9 M^{-1} s^{-1}. However, it is important to recognize that the stereospecific association of two proteins involves significant orientational constraints; thus, the Smoluchowski formula is of little use. Solc and Stockmayer tackled the problem of orientational constraints through a quasi-chemical approximation.[47] For two spheres each with a reactive patch they found the diffusion-controlled rate constant to be

$$k_{D0} = \frac{4\pi DRF_1F_2}{\Lambda_1\Lambda_2 + [(1-\Lambda_1)^{-1}(1-\Lambda_2)^{-1} + (1-\Lambda_1)^{-1}(\Lambda_2 - F_2)^{-1} + (1-\Lambda_2)^{-1}(\Lambda_1 - F_1)^{-1}]^{-1}} \tag{6}$$

where F_1 and F_2 are the surface fractions covered by the reactive patches. Relative diffusion has to bring their separation vector to within both reactive patches; therein the native complex can form instantaneously. Specifying the reactive patches thus amounts to specifying the transient complex. An approximate expression, obtained by Berg,[48] for Λ_i ($i = 1$ or 2), in the case of a patch spanning polar angles between 0 and δ_i is given by

$$\Lambda_i/F_i = \frac{\xi_i + \cot(\delta_i/2)}{\xi_i + \sin(\delta_i/2)\cos(\delta_i/2)} \tag{7}$$

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

$$k_{D0}/4\pi DR = F_1\xi_2 \tan(\delta_2/2) + F_2\xi_1 \tan(\delta_1/2) \tag{8}$$

The basal rate constant obtained by Brownian dynamics simulations (see section 2.2)[28] with simple structural assumptions about the transient complex (which was modeled by 2−3 correctly formed bonds between the proteins) is 10^5−10^6 M^{-1} s^{-1}, which is 4 orders of magnitude lower than the unrealistic Smoluchowski result. Such a rate constant is predicted by eq 8 for reactive patches at sizes of $\delta_1 \approx \delta_2 \approx 5°$ (with R and D at values quoted earlier and D_i at 0.02 ns^{-1}). Results similar to eq 8 for the basal rate constant have been obtained by a number of different methods.[49−52]

Reactive patches at sizes of ~5° cover only a fraction of 10^{-3} of the surface of each sphere. With the intersphere vector having to lie within both reactive patches before the native complex can form, the association rate constant is naïvely expected to be lower by a factor of $(10^3)^2 = 10^6$. As just stated, the reduction factor predicted by eq 8 is 10^4. The basal rate constant thus appears to be 100-fold higher than naïve expectation. Northrup and Erickson[28] explained this apparent paradox by noting, based on Brownian dynamics simulations, that when two proteins collide there is a high probability that they will recollide several times before separating again. Further, due to the separation of time scales between rotational and translational diffusion they may rotate significantly between those collisions.

Within the model of two spheres with reactive patches the influence of an interaction potential on the association rate constant has also been studied.[29,49] The expression for k_D in the presence of a centrosymmetric potential $U(r)$ is

$$k_D = \frac{4\pi DRe^{-U(R)/k_BT}[c_1(0)c_2(0)]^2}{\sum_{l,l_1,l_2=0}^{\infty}\left[\frac{\rho_l(r)/r}{-d\rho_l(r)/dr}\right]_{r=R}[c_1(l_1)c_2(l_2)]^2 C_{ll_1l_2}} \tag{9}$$

where $\rho_l(r)$ satisfies

$$r^{-2}e^{\beta U(r)}\frac{d}{dr}r^2e^{-\beta U(r)}\frac{d\rho_l(r)}{dr} - \left[\frac{l_1(l_1+1)D_1 + l_2(l_2+1)D_2}{D} + \frac{l(l+1)}{r^2}\right]\rho_l(r) = 0 \tag{10a}$$

with the boundary condition $\rho_l(r) \to 0$ as $r \to \infty$, $c_i(l_i)$ are given by Legendre polynomials

$$c_i(l_i) = P_{l_i-1}(\cos\delta_i) - P_{l_i+1}(\cos\delta_i) \qquad (10b)$$

and $C_{ll_1l_2}$ are given by Wigner $3-j$ symbols

$$C_{ll_1l_2} = \frac{2l+1}{(2l_1+1)(2l_2+1)}\begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} \qquad (10c)$$

Analytical results such as eq 9 can provide valuable insight on the properties of the association rate constant. In particular, in contrast to the Debye model in which rate enhancement by an attractive interaction potential is quite modest, the rate enhancement in the more realistic patch model can be substantial. For reactive patches with sizes of $\delta_1 \approx \delta_2 \approx 5°$, the rate enhancement is ~5000-fold for a potential with a magnitude of $9k_BT$ at contact.[29] More importantly, it is found that when $U(r)$ is long ranged and δ_1 and $\delta_2 \to 0$, the dependence of k_D on the interaction potential disappears except in the Boltzmann factor $e^{-U(R)/k_BT}$.[29,49] Therefore, the effect of the interaction potential can be simply captured by the Boltzmann factor. In section 2.2 we will return to this important result.

In the following sections we give a detailed account of theoretical and experimental studies of protein−protein association kinetics. In section 2 we outline the main experimental and theoretical methods for studying association kinetics. Section 3 presents experimental and theoretical analyses of the transient complex as well as the transition state for the final step in the association pathway. Section 4 continues with the important subject of electrostatic rate enhancement. In section 5 we discuss protein−protein association in the membrane environment, a field with growing interest. The subject of section 6 is the effect of crowding on association. Some final conclusions are drawn in section 7.

2. Methods for Studying Protein−Protein Association Kinetics

The association kinetics between proteins has been studied vigorously using both experimental and theoretical methods. There are some basic differences in studying binding of proteins and their folding. In folding the unfolded state is not well defined while the unbound state of proteins has a well-defined structure. This is not true in all cases as many protein−interaction domains were found in recent years to be unstructured, and thus, binding and folding in those cases are coupled.[53−56] However, here we review mostly the binding of structured proteins. A second main difference between binding and folding is that folding is a first-order process, where overall translational and rotational diffusion plays only a limited role. Conversely, as discussed in section 1, for binding overall translational and rotational diffusion play a major role in dictating the rate constant of the reaction. The existence of well-defined unbound structures simplifies the theoretical study of binding. For heterocomplexes this also simplifies the experimental study as one has a clear starting point for the reaction, the mixture of the separate components of the complex, and the possibility to follow in real time complex formation. In this review we focus on the experimental and theoretical work done to learn more about association between single globular proteins to form complexes.

2.1. Experimental Methods

Most of the experimental work reviewed here was done with the proteins being in free solution using methods such as stopped-flow spectrometry and NMR spectroscopy. This excludes the results of many measurements done using surface plasmon resonance (SPR) with one of the proteins attached to a surface and the other one in solution.[57] Direct comparisons of association rate constants for a number of protein−protein complexes showed that SPR data may provide different k_a values from those obtained in solution.[58,59] These differences may be attributed to mass transport, protein-immobilization effects, surface charges, crowding, etc.[60,61] As these factors introduce additional complexity to the analysis of protein association kinetics that is not directly related to the association reaction, we review here only solution data.

2.2. Computer Simulations and Modeling

One can model protein−protein association by simulating the translational and rotational Brownian motion of the subunits. In these Brownian dynamics simulations conformational fluctuations within the subunits are neglected, and hence, time scales far beyond those accessible to molecular dynamics simulations can be explored. Brownian dynamics simulations make it possible to calculate the diffusion-controlled rate constant k_D for protein shapes and interaction potentials beyond the scope of analytical theories. This approach has been used by many groups.[24,28,49,62−71] In such a calculation one must specify a precise set of conditions, which when satisfied signifies formation of the native complex. This set of conditions, typically implemented as an absorbing boundary in Brownian dynamics simulations, amounts to defining the transient complex. Rather than being guided by any theoretical considerations, the location of the absorbing boundary is usually proposed in an ad hoc way and often adjusted for best agreement with experiment. For example, in their Brownian dynamics simulations of barnase−barstar association Gabdoulline and Wade[24] tested three different specifications of an absorbing boundary against experimental data. Two of these were based on atom distances, and the third was based on electrostatic interaction energy; each specification had an adjustable parameter.

We will give a more detailed account on the specification of the transient complex in section 3.1. Here we address the effect of interaction potential, particularly one arising from electrostatic interactions between two associating proteins. In Brownian dynamics simulations the effect of electrostatic interactions can be modeled by accounting for their influence on the translational and rotational Brownian motion of the proteins. In principle, the electrostatic force and torque on the proteins can be calculated by solving the Poisson−Boltzmann equation

$$\nabla\varepsilon(r)\,\nabla\,\Phi(r) = -4\pi\left[\rho(r) + M(r)\sum_{i:\text{ions}} c_i q_i e^{-q_i\Phi(r)/k_BT}\right] \qquad (11)$$

where ε is the dielectric constant, Φ is the electrostatic potential, ρ is the charge density inside the proteins, M is a function with a value of 1 in all regions where mobile ions in the solvent are accessible and 0 elsewhere, and q_i and c_i are the charges and bulk concentrations of mobile ions, respectively. The solution of the Poisson−Boltzmann equa-

tion is simplified by linearization leading to

$$\nabla\varepsilon(r)\nabla\Phi(r) = -4\pi\rho(r) + M(r)\varepsilon_s\kappa^2\Phi(r) \qquad (12)$$

where ε_s is the dielectric constant of the solvent and $\kappa = (8\pi e^2 I/\varepsilon_s k_B T)^{1/2}$ is the Debye−Hückel screening parameter with $I = \sum_{i:\ \text{ions}} c_i q_i^2 /2e^2$, the ionic strength. However, even after linearization, solving the Poisson−Boltzmann equation on the fly during a Brownian dynamics simulation is prohibitively expensive. One thus has had to rely on approximations such as treating one of the proteins as a set of test charges[62] (which leads to significant errors from neglecting the low-dielectric region of the protein interior[49]) or the more elaborate effective-charge model.[72] Unfortunately, the approximations are worst when the proteins are in close proximity, precisely where electrostatic interactions are expected to have the strongest influence on k_D.

An algorithm is needed to obtain k_D from the simulations. A widely used algorithm was developed by McCammon and co-workers.[73] According to this algorithm one protein is translationally immobilized at the origin and trajectories of the other protein are initiated uniformly on a spherical surface (at a radial distance $r = b$) (see Figure 2a). This "b" surface should be chosen far away from the immobilized protein such that the electrostatic interaction potential is centrosymmetric. If rotational motion is allowed, then the starting orientations of the two proteins are random. The diffusion-controlled rate constant is given by

$$k_D = 4\pi D b f_\infty \qquad (13)$$

where f_∞ is the fraction of trajectories that, instead of escaping to infinity, reach the absorbing boundary specifying the conditions for formation of the native complex. To find f_∞, an outer absorbing boundary at $r = q > b$, called the q surface, is introduced. The trajectories are terminated when either the absorbing boundary for complex formation or the q surface is reached. From the fraction, f, of trajectories terminated on the inner absorbing boundary one finds

$$f_\infty = \frac{f}{1 - (1-f)k_D(b)/k_D(q)} \qquad (14)$$

where $k_D(b)$ and $k_D(q)$ are given by eq 5 with R replaced by b and q, respectively. It has been shown that eq 14 is strictly valid only when the q surface is far away from the immobilized protein such that the equilibrium distribution of the mobile protein becomes isotropic.[74]

An alternative algorithm[75,76] has also been developed in which the absorbing boundary is extended into a "reaction region" with a finite volume V_{RR} (Figure 2b).[18,77] A protein pair that has reached the reaction region has a finite rate, γ, to form the native complex. In this treatment of protein association the reaction rate γ models the conformational rearrangement that brings the protein pair from the transient complex (specified by the reaction region) into the native complex, that is, $k_c = \gamma$. The equilibrium constant k_D/k_{-D} is given by $V_{RR}e^{-\langle U\rangle^*/kBT}$, where $\langle U\rangle^*$ is the average interaction energy within the transient complex. By starting Brownian trajectories from within the reaction region and obtaining the surviving fraction, S, of the trajectories, the association rate constant is obtained as $k_a = \gamma S V_{RR} e^{-\langle U\rangle^*/k_B T}$. Substituting the various results into eq 2, one finds that the diffusion-controlled association rate constant is given by

$$k_D = \gamma V_{RR} e^{-\langle U\rangle^*/k_B T} S/(1-S) \qquad (15)$$

While the surviving fraction S depends on the rate γ and on how much the absorbing boundary is extended to form the reaction region, the result for k_D given by eq 15 is only determined by the absorbing boundary and the interaction energy. We also note that the reaction-controlled association rate constant (eq 3b) is

$$k_R = k_c k_D/k_{-D} = \gamma V_{RR} e^{-\langle U\rangle^*/k_B T} \qquad (16)$$

in the present treatment.

In a Brownian dynamics study of protein−protein association under the influence of electrostatic interactions it was discovered that the survival fraction S is insensitive to the presence of the electrostatic interaction energy.[49] Then one can write eq 15 as

$$k_D = k_{D0} e^{-\langle U\rangle^*/k_B T} \qquad (17)$$

with k_{D0} as the basal rate constant given by

$$k_{D0} = \gamma V_{RR} S_0/(1-S_0) \qquad (18)$$

where S_0 is the survival fraction without any biasing force. Note that a dependence on interaction energy as simple as that given by eq 17 is expected for the reaction-controlled rate constant (see eq 16), but totally unexpected for the diffusion-controlled rate constant. Later an analytical derivation of eq 17 was presented,[78] which confirmed the two requirements for the validity of eq 17 suggested in the original work.[49] These requirements are that the association is stereospecific and the interaction energy is long ranged. Fortuitously (and fortunately!) these conditions are fulfilled by protein−protein association under the influence of electrostatic interactions. The accuracy of eq 17 has been demonstrated against results from Brownian dynamics simulations of protein−ligand binding[20,79,80] and against

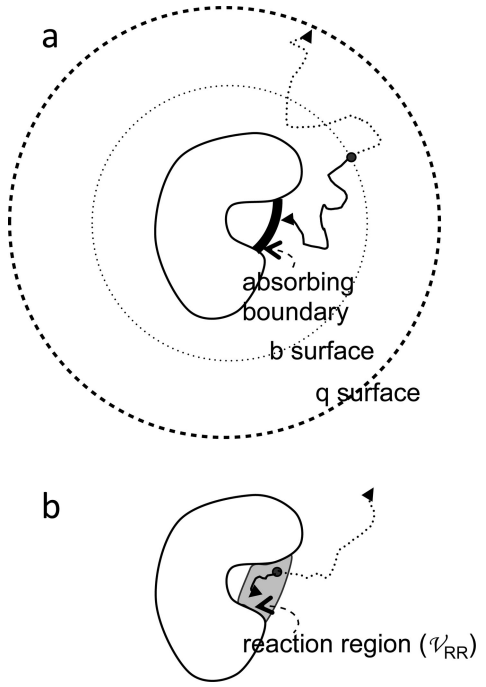


Figure 2. Two algorithms for calculating the diffusion-controlled association rate from Brownian dynamics simulations. (a) Algorithm of Northrup et al.[73] (b) Algorithm of Zhou.[75] In either algorithm Brownian trajectories can be terminated by two triggers. In a the triggers are absorption by either the inner absorbing boundary or the q surface; in b the triggers are either reaction in the reaction region or when the lifetime of the trajectory exceeds a cutoff.

analytical results given by eq 9 for model systems.[29] This equation resolves one of the two main obstacles to reliable prediction of protein association rate constants by making it possible to rigorously treat electrostatic interactions. The effect of electrostatic interactions is captured by the Boltzmann factor $e^{-\langle U\rangle^*/k_BT}$, which can be obtained by averaging over a relatively small number of representative configurations in the transient complex. The basal rate constant k_{D0} still needs to be obtained through force-free Brownian dynamics simulations, but these simulations are inexpensive.

At first glance eq 17 looks like the transition-state theory of the Eyring type.[25] There is no relation between the two. As already emphasized, eq 17 is for a diffusion-controlled rate constant, but Eyring's theory is for energy barrier-crossing processes that are activation controlled. Kramers[27] has shown that Eyring's theory is a very poor approximation when barrier crossing becomes diffusion controlled. However, there is no link between eq 17 and Kramers' theory for diffusion-controlled barrier crossing either. In particular, the average interaction energy of the transient complex, $\langle U\rangle^*$, which superficially resembles the energy barrier in Eyring's and Kramers' theories, is typically negative (leading to rate enhancement) instead of positive. Unlike the processes studied by Eyring and Kramers, where the energy barrier is what hinders the rate, the diffusion-controlled rate of protein association is hindered by the translational/rotational constraints of the transient complex. Moreover, eq 17 does not require a reaction coordinate like the theories of Eyring and Kramers. In fact, as will be discussed next, the specification of the transient complex involves at least six degrees of freedom: three for translation and three for rotation.

3. Intermediates along the Association Pathway

Throughout this review we use the word "intermediate" as a generic term for a set of configurations in which there is a degree of correlation in translation and rotation between the associating proteins; it does not necessarily correspond to a local energy minimum or a species trapped in kinetic experiments. We refer to the latter as a kinetic intermediate. As discussed in section 1.1, the association process starts with formation of the transient complex (sometimes termed the encounter complex, see section 3.1 below). The transient complex is close to the native complex in relative separation and relative orientation between the subunits but still misses a majority of the short-range interactions characterizing the native complex. Formation of these short-range interactions requires passage through another fleeting structure, the transition state for conformational rearrangement.[26,81] How can structural information be obtained on these intermediates?

Structural studies are routinely done to elucidate the transition state and kinetic intermediates of protein folding or enzyme catalysis. A range of experimental tools has been developed for this task. NMR is a powerful tool to pin down the residual structures of the unfolded state as well as capture transient folding intermediates.[82] Φ-Value analysis reveals whether specific interactions are formed already during a kinetic intermediate or the transition state of the reaction.[56,83,84] Time-resolved spectroscopy and single-molecule spectroscopy are powerful tools which are frequently applied to investigate kinetic intermediates and transition states in folding.[85−88] While these experimental tools provide only a partial view, they are extremely valuable for molecular dynamics simulations and other computational studies as they provide experimental reference points to benchmark simulations.[89−91]

In comparison to studies on protein folding, structural studies on intermediates along the pathway of protein−protein association are much less common. This may be partly attributed to the technical difficulties stemming from the low population of the binding intermediates and the ill-defined nature of the transition state for binding. Still, the development of protein-engineering tools, NMR, and time-resolved optical spectroscopy resulted in a number of interesting experimental studies shedding light on the mechanisms of protein−protein association. Valuable guidance has also been provided by theory. In the following we focus on studies characterizing the transient complex and the transition state of association.

3.1. Transient Complex

In theoretical predictions of association rate constants one must specify the transient complex. In different contexts the transient complex has appeared as reactive patch, absorbing boundary, reaction region, and encounter complex. The ad hoc way by which the transient complex is specified in Brownian dynamics studies is pointed out in section 2.2. The situation was no better when eq 17 was first used to predict association rate constants. In one early such study on barnase−barstar association[77] the transient-complex ensemble was specified by adjusting the ranges of relative translation and rotation between the two proteins to match the experimental data at high ionic strength. Miyashita et al.[92] directly used experimental data for the association of cytochrome $c2$ with the bacterial reaction center to locate the transient-complex ensemble in the 6-dimensional translation and rotation space. The correlation coefficient between calculated changes in the electrostatic interaction energy by mutations and the experimental counterpart

$$\Delta\Delta G^{\ddagger-U} \equiv -k_BT \ln(k_a^{mut}/k_a^{wt}) \tag{19}$$

was used as an effective energy in a Monte Carlo simulation. Schreiber and co-workers[93,94] further simplified eq 17 by calculating $\langle U\rangle^*$ from an empirical function simply on the native complex, thus avoiding the specification of the transient complex; the empirical function was parametrized on experimental data. Applications of that approach are presented in section 4.2.

In order to predict association rate constants from theory alone, the transient-complex ensemble has to be specified without reference to experiment. A solution to this challenging problem was proposed in a recent paper[18] based on analyzing the interaction energy landscape of associating proteins.

In a complete theory the overall association rate constant k_a should not be sensitive to where the transient complex is placed. If it is placed far away from the native complex, then k_D will be large but k_c will be small. Conversely, if it is placed very close to the native complex, then k_D will be reduced but k_c will become very large. Either way, eq 2 is expected to give nearly the same result for k_a (as explicitly shown for a model system[95]). However, given the considerable difficulty and uncertainty in the calculation of k_c it is highly desirable to use k_D as a close approximation for k_a. Then there is an optimal location for placing the transient complex.[22] If it is placed too far from the native complex,

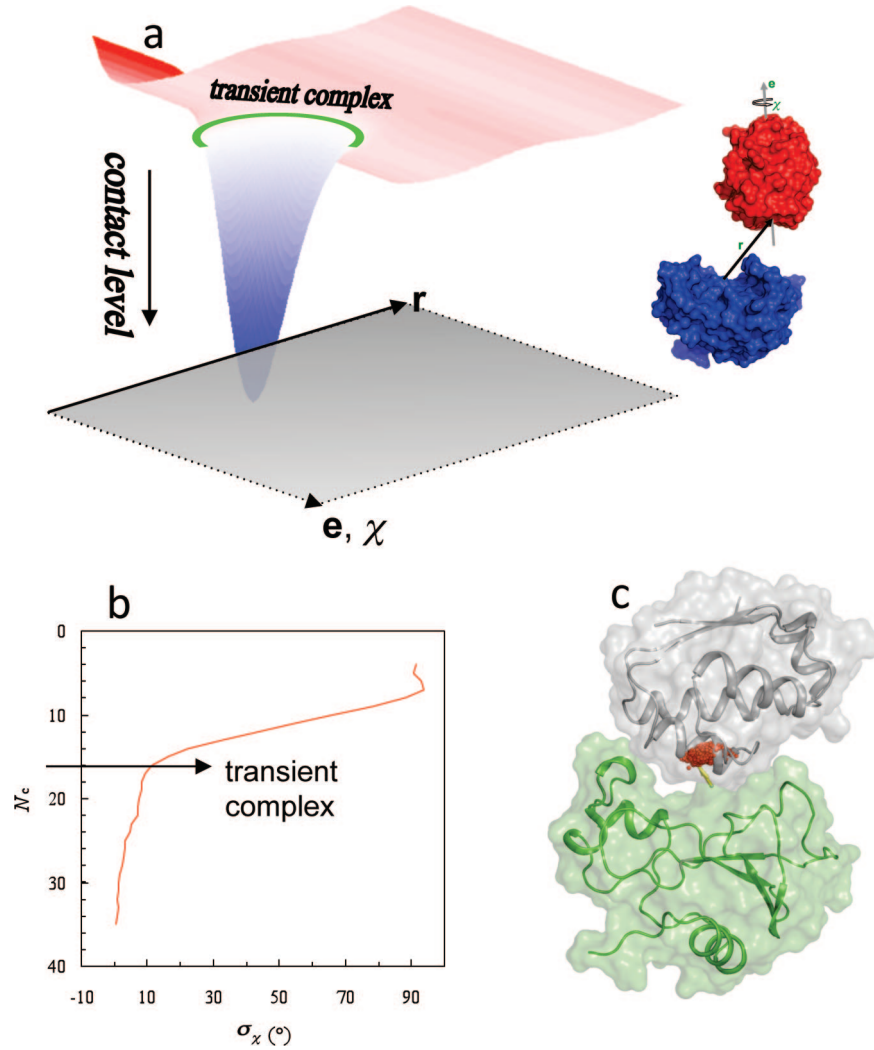


Figure 3. Location of the transient complex within the interaction energy landscape. (a) The energy landscape of protein−protein association. Translation (**r**) and rotation (**e**, χ) coordinates lie in the horizontal plane. The transient−complex ensemble is indicated by a green ring. The smooth energy landscape results from keeping the two proteins in their native internal conformations. (b) Identification of the transient complex, consisting of configurations with the contact level N_c*, where σ_χ is poised for a sharp increase with decreasing N_c. (c) Translational volume of the transient complex, which is the projection of the hypersurface $N_c = N_c$* into the 3 dimensions of translation, for the barnase−barstar pair. A cluster of red dots represent the locations of the displacement vector **r** (yellow arrow). (Parts a and b were adapted with permission from ref 1. Copyright 2008 Wiley Interscience.)

then the resulting k_D would not be a useful approximation for k_a. On the other hand, placing the transient complex too close to the native complex would mean that short-range interactions and conformational rearrangement have to be dealt with in calculating k_D. The native complex sits in a deep well in the interaction energy landscape.[18] The optimal placement for the transient-complex ensemble is at the outer boundary of the native-complex energy well (Figure 3a).[18,22]

The algorithm for identifying the transient complex was based on the following observation: inside the native-complex energy well translation and rotation are restricted, but once outside, the proteins gain significant translational/rotational freedom.[18] Thus, the outer boundary of the native-complex energy well coincides with the onset of translational/rotational freedom. To simplify the calculations required for determining the transient complex, the short-range interaction energy stabilizing the native complex was modeled by the number of contacts, N_c, formed between the protein partners. Translational/rotational freedom was measured by $\sigma_\chi(N_c)$, the standard deviation of the rotation angle χ in configurations with a given contact level N_c. A sharp increase in σ_χ with decreasing N_c marks the onset of translational and rotational freedom and hence the location of the transition complex (Figure 3b). As illustrated in Figure 3c for the barnase−barstar pair, the resulting model for the transient complex is a compact ensemble of configurations which marks the onset of a sharp increase in translational and rotational freedom as the protein partners move out of the native-complex energy well. In the ensemble of transient-complex configurations of the barnase−barstar pair the displacement vector **r** is mostly confined in a $4 \times 4 \times 1$ Å^3 volume, corresponding to distances of 4.5 ± 0.5 Å between the protein surfaces (allowing for at least one layer of solvent). Within this translational volume the relative orientations of the two proteins are also severely restricted with the rotation angle χ mostly confined in the 0°−10° range.

Since k_D is used as the prediction for k_a, eq 17 can be rewritten as an equation for k_a

$$k_a = k_{a0} e^{-\langle U_{el}\rangle^*/k_B T} \tag{20}$$

where k_{a0} is the rate constant for reaching the transient complex in the absence of any biasing force and can be obtained by Brownian dynamics simulations, and the interaction energy is now denoted with a subscript "el" to signify

that only electrostatic interactions are to be included. The neglect of short-ranged nonelectrostatic effects from the Boltzmann factor is based on two considerations. First, as Figure 3c shows, the transient-complex configurations are separated by at least one layer of solvent; therefore, short-ranged forces such as hydrophobic and van der Waals interactions are relatively weak in the diffusion process, leading to the transient complex. Second, short-range interactions, even when present within the transient complex, contribute much less to rate enhancement (i.e., k_a/k_{a0}) compared to long-range interactions. However, short-range interactions are essential for determining the location and size of the transient-complex ensemble in configurational space, which in turn affect the magnitude of k_{a0}. A transient-complex ensemble that is less restricted in translation and rotation will lead to a higher k_{a0}. Variation of the restriction in translation and rotation within the transient complex with solvent conditions or among different protein complexes can be viewed as a configurational entropy effect. The basal rate constant k_{a0} captures all the contributions of short-range interactions and configurational entropy.

The structural model for the transient-complex ensemble presented above along with eq 20 constitutes the transient-complex theory for predicting protein−protein association rate constants. With this theory the two main obstacles faced by the traditional approach of Brownian dynamics simulations, the necessity for approximate treatment of electrostatic interactions and the ad hoc specification of the transient complex, are both resolved. Applications of this theory are presented in section 4.3.

3.2. Other Intermediates

The term "encounter complex" has frequently been used in describing the pathway of protein−protein association. It appears that this term is assigned different meanings in different contexts. For example, in calculating association rate constants by Brownian dynamics simulations, Gabdoulline and Wade used this term to refer to "the end-point of diffusional association",[24,96] which would be similar to what we have defined as the transient complex. However, these same authors later also used encounter complex to refer to low free-energy regions in configurational space.[70] In the paramagnetic NMR experiments described next, encounter complex refers to a minor, dynamic state that is in equilibrium with a dominant, stereospecific complex.

Using the paramagnetic relaxation enhancement (PRE) NMR technique on the electron-transfer pair of yeast cytochrome *c* peroxidase (CcP) and iso-1-cytochrome *c* (*Cc*) Volkov et al.[97] determined the structure of the dominant, well-defined complex and delineated the configurational space of the minor, dynamic state. The latter, as just noted, was referred to as encounter complex. The dominant complex is very similar to the X-ray structure of the pair and occupied for >70% of the time. In the encounter complex, *Cc* occupies a region around the position in the dominant complex; the negative charges of CcP bordering this region suggest "that electrostatic attraction plays a dominant role in determining the nature of the encounter complex". However, the data also seem to indicate that the encounter complex contains configurations in which *Cc* has its "back" side facing CcP.

Tang et al. carried out a similar study for the association between the phosphocarrier protein, Hpr, and three proteins in the bacterial phosphotransferase system.[98] They found a rather diffuse encounter complex that qualitatively correlates with the electrostatic surface potentials on the interacting proteins. These experiments were performed in the absence of salt, in which case nonspecific electrostatic attraction of these highly charged proteins is strong over a long range.[6] In a follow-up paper[99] Tang et al. showed that the nonspecific part of the encounter complex is reduced to a large extent by adding salt, while more specific configurations (located near the region of the native complex) were less affected. This study seems to suggest that not all regions of the encounter complex lead to productive association; it is the region which does lead to productive association that we specifically refer to as the transient complex.

Spaar et al.[70] developed a method for mapping the encounter complex along the association path from BD simulations. Applying this method to the barnase−barstar interaction they suggest two minima along the pathway, one leading (according to an analysis of successful trajectories) to association, while the other does not. Still, they suggest that this second region may be helpful to steer barstar into the region of the transient complex. In a yet unpublished study combining simulation and experiment Spaar et al. aimed to directly evaluate the importance of nonspecific encounter regions to association. Using the method developed by Spaar et al.[70] on TEM1−BLIP association, the BD simulations identified two main regions with low interaction energies for the wild-type complex; however, most trajectories started at these regions did not lead to productive binding. It was found that only mutations designed to increase the size and energy of the encounter region near the binding interface had a major effect on the measured association rate constant. Conversely, mutations increasing or decreasing the occupancy and energy of other intermediates had no effect on the association rate constant. These results indicate that not only the occupancy or energy of the encounter regions but also their location are important in determining their effect on the association rate constant. Thus, some encounter regions are fruitful while others are futile. This conclusion was in fact anticipated when the encounter region that does lead to productive association was singled out to define the transient complex (see above discussion illustrated by Figure 3a).[1]

In some cases, kinetic intermediates in binding have been observed. One such case is the interaction between colicin endonucleases and their cognate and noncognate immunoproteins.[100] Stopped-flow spectrofluorimetry shows that binding of both Im9 and Im2 to E9 result in a rapid fluorescence quenching step (pre-equilibrium) with a concentration-dependent rate of formation followed by a second, concentration-independent slow rate (on a time scale of seconds). The data are best explained by a mechanism where cognate and noncognate complexes alike form a "dynamic encounter complex". Thereafter the cognate pair transitions to a high-affinity, stereospecific complex, whereas the noncognate pair remains dynamic without achieving high affinity.

A somewhat different mechanism involving a pre-equilibrium complex followed by a reorganization step was suggested for cystatin A binding papain.[101,102] The biphasic behavior of cystatin A/papain interaction was observed by following independently the inhibition of the catalytic activity of papain (where a linear dependence of k_{obs} on protein concentration was observed), while the spectroscopic probe (which monitors the accumulation of the native complex) showed a hyperbolic dependence of k_{obs} on protein concen-

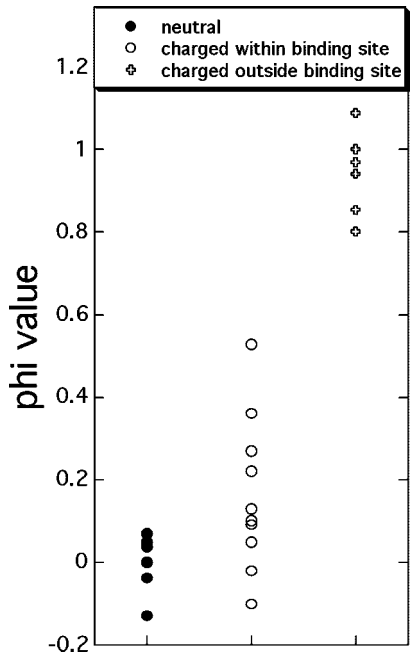


Figure 4. Φ-Value analysis of the transient complex/transition state for association determined for 'hotspot' mutations (affecting the binding affinity by >2 kcal/mol) collected from TEM1−BLIP, barnase−barstar, and Ras−Ral association.[6,14,93] Φ values close to 1 suggest that the involved residues have similar interactions in the transient complex/transition state and in the native complex, while values close to 0 indicate the residues do not form any interprotein contacts in the transient complex/transition state.

tration. This example demonstrates that characterization of the association behavior may depend on the probe used to monitor it.

3.3. Mapping Intermediates by Mutations

3.3.1. Φ-Value Analysis

Φ-Value analysis was successfully applied to map the transition state for protein folding and became the golden standard for many simulations.[103] It has been demonstrated that this analysis can be used also for studying the rate-determining state (denoted by ‡) for protein−protein association.[14,56,104−106] The basic formulation for such analysis for binding is given by

$$\Phi_a = \Delta\Delta G^{\ddagger-U}/\Delta\Delta G^{C-U} \quad (21)$$

where $\Delta\Delta G^{\ddagger-U}$ is defined by eq 19 and $\Delta\Delta G^{C-U}$ is analogously defined but with the association rate constant k_a replaced by the association constant K_a. Mutations that induce a similar effect on the rate-determining state (‡) and the free energy of binding will have a Φ value of 1, while mutations that have no effect on the rate-determining state (‡) but change the association constant will have a Φ value of zero. The interpretation of this form of Φ-value analysis is particularly transparent when conformational rearrangement is rate limiting (cf. eq 3b). In this case $\Delta\Delta G^{\ddagger-U}$ represents the mutational effect on the free energy for forming the transition state between the transient complex and the native complex, and the Φ-value analysis thus reports on the transition state. If, on the other hand, conformational rearrangement is fast (i.e., $k_c \gg k_{-D}$) then the analysis reports on formation of the transient complex.

Figure 4 shows a Φ-value analysis for a large number of 'hotspot' mutations (those with $\Delta\Delta G^{C-U} > 2$ kcal/mol in magnitude) collected from TEM1−BLIP, barnase−barstar, and Ras−Ral binding.[6,14,93,107] The mutations were divided into three groups: one consisting of noncharged residues, the second consisting of charged residues located within the binding interface, and the third consisting of charged residues located outside the binding interface. The reason we show only 'hotspot' residues is to avoid erroneous Φ values as the experimental error for $\Delta\Delta G$ measurements is on the order of 0.3 kcal/mol. As only very few residues located outside the binding interface pass this criteria, we included in this group also multiple mutations that were designed to specifically increase the association constant.[14,93] The data clearly demonstrate that noncharged mutations always have Φ values close to zero, while charged residues located outside the physical binding site have Φ values close to one. Charged residues located within the binding site have mixed values. A simple explanation of these Φ values is that they report on the transient complexes of the three protein pairs. As noted in the discussion below eq 20, the transient complex is predominantly stabilized by long-range electrostatic interactions with nonpolar interactions playing a minor role. A striking result of the Φ-value analysis is that hydrophobic residues appear to play only a minor role in formation of the transient complex. As most mutations were made to Ala, which retains a hydrophobic surface, one cannot rule out the possibility that nonspecific hydrophobic interactions are formed at the transient complex also for the mutant, reducing the Φ value. However, this contribution should be very limited as mutations to Ala create cavities on the surface that are not fully filled up by the incoming protein.

A hallmark of a partially reaction-limited reaction (i.e., not purely diffusion controlled) is a nonlinear relation between the rate of association and the concentrations of reactants. This has been observed for enzyme−substrate interactions but rarely for protein−protein association. However, in the case of the association between RalGDS-RBD and Ras this nonlinearity could be experimentally demonstrated, presenting direct evidence for the transient complex as a kinetic intermediate for association.[14] The nonlinearity was observed despite a high k_a value (10^7 M^{-1} s^{-1}) and a putative basal rate of 6×10^5 M^{-1} s^{-1}, which should in principle place this association in the diffusion-controlled regime (see section 1.1). Using the program PARE (see section 4.2.1),[93] mutations of charged residues located at the periphery of the binding interface were designed, which specifically increased k_a and did not affect k_d. Stopped-flow measurements showed that the increase in k_a of a +7 mutant was a result of an increased rate of formation of the transient complex, while the rate of conversion to the native complex was unchanged at a value of ~400 s^{-1}.[14] From these data Φ values of 1.0 and 0.9 for the transient complex and transition state, respectively, were calculated. This demonstrates that favorable electrostatic interactions introduced by mutation stabilize the transient complex and transition state to the same extent as they stabilize the native complex. This study also points to the difficulty in assigning a particular association process as diffusion limited or reaction limited. At low protein concentrations the overall association rate may appear to be diffusion limited, but at high protein concentrations it may appear to be reaction limited.

The +7 RalGDS−RBD variant (obtained by electrostatic design) binds Ras with an association constant similar to that measured for the native effector, Raf. Interestingly, the

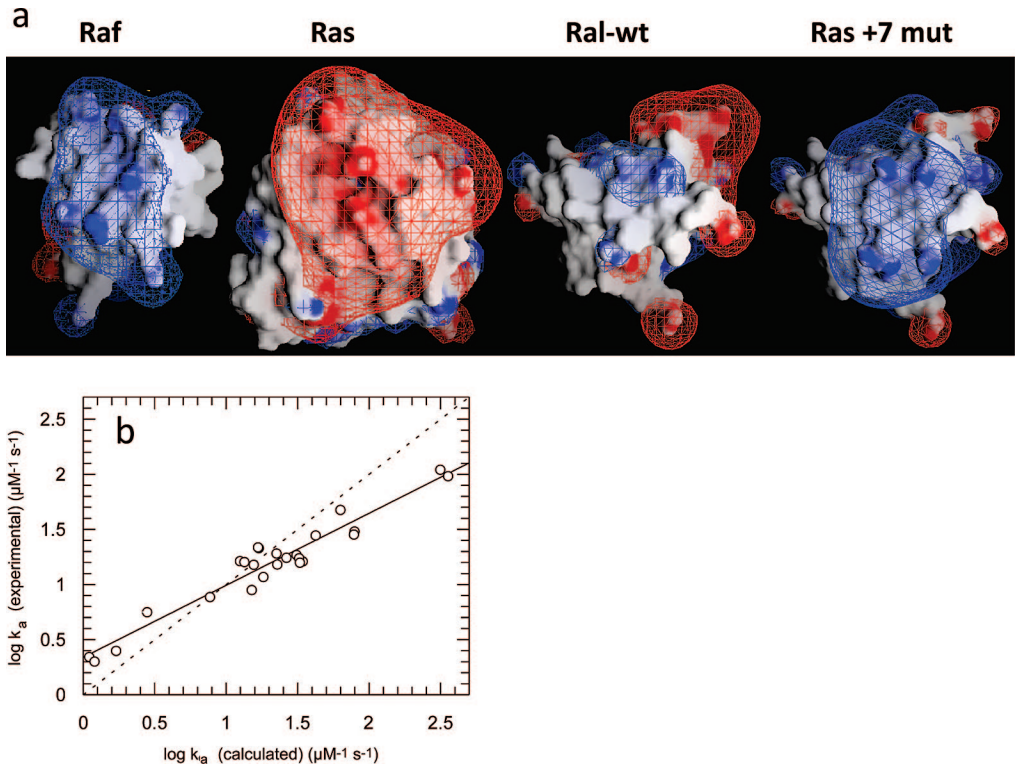


Figure 5. (a) Electrostatic potential maps of Ras, Raf−RBD, and Ral−RBD. The Ral + 7 mutant contains the mutations M26K/D47K/E54K/D90K, which were predicted by the program PARE to significantly increase the rate constant of association of Ral−RBD to Ras. These four mutations are located outside the Ras-binding site on Ral−RBD. The figure was drawn using the program GRASP with the contours drawn at $2k_BT/e$ (blue for positive and red for negative). (b) Plot of experimental values of the association rate constants for mutants of Ral−RBD to Ras vs values calculated using PARE. (Adapted with permission from ref 14. Copyright 2004 The National Academy of Sciences of the United States of America.)

electrostatic potential map of the RalGDS−RBD variant is also similar to that of Raf (Figure 5a) despite the very different sequences of the two effectors (<15% homology).[14] The initial aim of this project was to optimize through mutation the electrostatic interactions between RalGDS-RBD and Ras. The similarity of the electrostatic potential maps of the +7 Ral mutant and Raf suggests that the natural complex between Ras and Raf is optimized by natural selection for fast binding.

pK_a shifts during the association process can also provide structural information about the transient complex.[108,109] The pK_a of His102 in unbound barnase was measured to be 6.3, while in the native complex a pK_a of <5 was measured. The pH dependence of k_a showed a pK_a value similar to that in the unbound protein; therefore, the shift in pK_a upon binding barstar occurs after the transient complex. The X-ray structure of the native complex and electrostatic calculations[110] indicate that the shift in pK_a is attributable to the burial of His102 in the interface with barstar and interactions with surrounding barstar residues; the pH dependence of k_a suggests that the burial and interactions are not fully formed within the transient complex. A similar behavior was observed for the association of R67 DHFR with a pK_a of 6.6 that was assigned to H62 but a dissociation reaction with a pK_a of under 5.5 with the pK_a shift being attributed to specific short-range interactions that are not formed within the transient complex/transition state for association.[111] These studies provide a clear indication that short-range interactions are mostly formed late along the association pathway.

3.3.2. Double-Mutant Cycle Analysis

Double-mutant cycles measure the coupling energy between a pair of residues from the difference in binding free energy between a double mutant and the two single mutations. For binding, the coupling energy defined by[108,112−114]

$$\Delta\Delta G_{int}^{\ddagger-U} = \Delta\Delta G_{X\rightarrow A,Y\rightarrow A}^{\ddagger-U} - \Delta\Delta G_{X\rightarrow A}^{\ddagger-U} - \Delta\Delta G_{Y\rightarrow A}^{\ddagger-U} \quad (22)$$

provides a measure of the interaction between two residues, X and Y, in the transient complex or the transition state (depending on which dictates the rate-limiting step). Like Φ-value analysis, this technique was also first applied successfully to protein-folding studies.[112,113,115,116]

From a large number of double-mutant cycles studied concerning the association rate constant of barnase and barstar significant coupling energies were found only between charged residues separated by <10 Å from one another in the native complex.[6] None of the noncharged residues had a significant $\Delta\Delta G_{int}^{\ddagger-U}$ value with any other residue. A similar experiment was done on the interaction between cytochrome C_2 and the bacterial reaction center, but only between charged residues. $\Delta\Delta G_{int}^{\ddagger-U}$ values, as calculated from the measured electron-transfer rate constant, which is closely related to k_a,[92] again showed that significant interactions exist in the transient complex/transition state only between residues which are <10 Å away from each other in the native complex.[117] A similar result was obtained also for association of TEM1−β-lactamase and its protein inhibitor BLIP.[118] Repeating the double-mutant cycles at up to 1 M salt (which masks the effects of charge−charge interactions) showed that

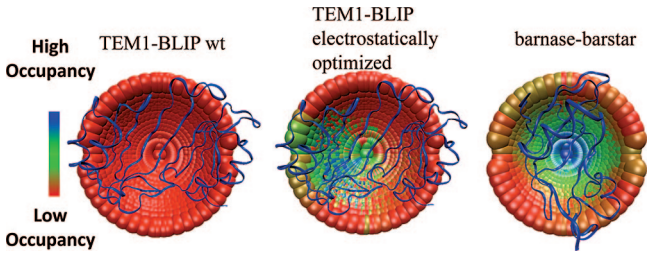


Figure 6. Mapping the transient complex/transition state for protein−protein association for the TEM1−BLIP and barnase−barstar complexes using double-mutant cycle data as constraints. Each point represents the center of mass of one of the 2220 configurations of TEM1 or barnase perturbed from the native complex (the mobile proteins in the simulations). BLIP and barstar are the fixed proteins and are represented as ribbon. The point in the middle of each cap represents the X-ray structure of the native complex. The different colors represent configurations selected by different filtering cutoffs; colder colors designate a configuration that passes a more stringent cutoff (thus, has a higher probability of occupancy in the transient complex/transition state). The TEM1−BLIP complex was electrostatically optimized using the program PARE by introducing mutations located outside the binding interface.

for barnase−barstar and the complex between *P. laminosum* Cyt*f* and plastocyanin some but not all pairwise charge−charge interactions were maintained, suggesting that the structural specificity of the transient complex/transition state is preserved even at high salt but its size may shrink.[119,120]

The experimental mutant and double-mutant cycle data measured on the association rate constants were further used to model the structures of the transient complex/transition state. In the study of Harel et al. the transient complex/transition state was modeled from the experimental $\Delta\Delta G^{\ddagger-U}_{int}$ values by searching for those interprotein orientations that best account for the experimental $\Delta\Delta G^{\ddagger}_{int}$ values (Figure 6). Similarly, Miyashita[92] related the experimental k_a values of mutant proteins to differences in the calculated electrostatic energies for a wide range of cytochrome C_2−reaction center (Cyt−RC) configurations. Both studies gave a very similar description of the transient complex/transition state. In both cases, the transient complex/transition state was stabilized by electrostatic interactions with the ensemble of configurations spread out around the native complex, but in neither case were short-range interactions formed, suggesting a solvated transient complex/transition state. The average structure of the transient complex/transition state was not necessarily located exactly at the center of the binding interface but could be shifted toward one side of the interface. This was observed for both the Cyt−RC and the electrostatically optimized TEM1−BLIP pairs, while for the barnase−barster pair the average structure of the transient complex/transition state overlapped the native complex.[92,94,118] These characteristics bear striking resemblance to the compact structural model of the transient complex derived from theory (section 3.1). They suggest a certain pathway for association (down an energy funnel), which would help in speeding up the association.[18] Contrary to these results, no indication for a transient complex with specific structures was found for the association between wild-type TEM1 and BLIP or between IFNα2 and IFNAR2 (Figure 6). Therefore, a diffusive transient complex/transition state was suggested for these protein pairs. What distinguishes a diffusive transient complex from a compact one is the absence of defined interprotein orientations. As was clearly shown for the TEM1−BLIP complex, mutations introduced through rational design can change the transient complex/transition state from diffusive to compact and vice versa.[118]

All the evidence described above suggest the existence of a transient complex/transition state along the association pathway. The main points presented are (1) the binding of Ras to either Ral or Raf is biphasic with the first phase being diffusion limited and the second phase reaction limited, (2) increasing the electrostatic complementarity between Ras and Ral specifically affects the rate of the first phase and does not affect the second phase of association, (3) increasing electrostatic complementarity affects only the association but not dissociation rate constant, (4) double-mutant cycle experiments as well as Φ-value analyses clearly show that only charged residue pairs form interactions in the transient complex, while short-range interactions are largely unformed in the transient complex. Further formation of such short-range interactions, which are important for stabilizing the native complex, may encounter a transition state. In the case of a compact transient complex the transient complex already has interprotein orientations similar to those in the native complex. The transition state will at least inherit such an orientational similarity. Therefore, in terms of gross structure, the transient complex and transition state seem to be close.

4. Probing Electrostatic Enhancement of Association Rates

As discussed in previous sections, electrostatic interactions have long been recognized as the dominant factor used by Nature in order to enhance protein association beyond the basal rates dictated by diffusion.[13,49,121] In this section we discuss experimental and theoretical methods to probe and analyze the effect of electrostatic attraction on association kinetics as well as protein engineering studies which enable us to modify and control this effect. The preceding sections have already presented a glimpse into some of these issues.

4.1. Ionic Strength Dependences of k_a and k_d

The magnitude of the electrostatic attraction between two proteins can be modulated most simply by changing the ionic strength of the solution. It has been recognized that when the association between two proteins is diffusion controlled modulation of the electrostatic attraction leads to a universal phenomenon related to the association and dissociation rate constants k_a and k_d, which show very disparate dependences on ionic strength.[22,122] The association rate constant decreases significantly with increasing ionic strength, whereas the dissociation rate constant is only modestly affected by ionic strength. The disparate ionic strength effects have been observed on a number of the protein−protein complexes discussed above.[4,5,11,35,109] Many other proteins conform to the same behavior.[123−133]

The compact structural model of the transient complex derived from theory, presented in section 3.1, provides a nice explanation for the disparate effects of ionic strength on association and dissociation rate constants when association is diffusion controlled.[22,122] As the transient complex lies at the outer boundary of the interaction energy well and hence is close to the native complex, ionic strength is expected to screen electrostatic interactions in the two types of complexes to nearly the same extent. Hence, the association constant and association rate constant are expected to have nearly the same dependence on ionic strength and the dissociation rate would be little affected by ionic strength.

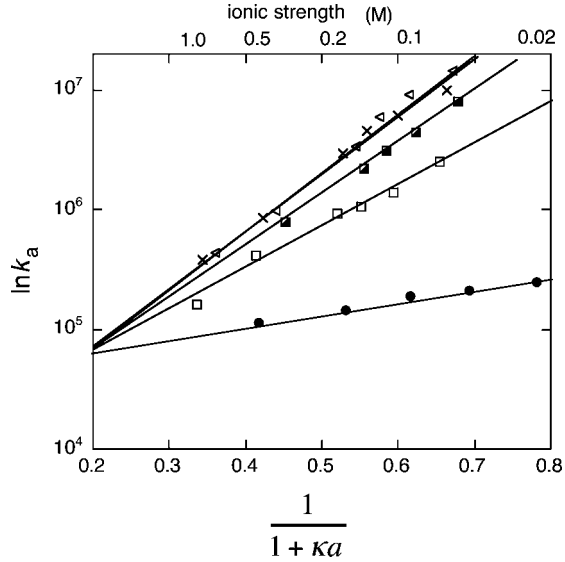


Figure 7. Association rate constants of wild-type and mutant TEM1−BLIP complexes determined at different salt concentrations with ln k_a plotted against $(1 + \kappa a)^{-1}$. The data can be fitted to a line according to eq 23.

The relation between ionic strength and k_a can be quantitatively predicted from the solution of the Poisson−Boltzmann equation (eq 12) as will be discussed in section 4.3. Empirically, it has been shown that this relation can be described quite well by a Debye−Hückel-like approximation[77,93]

$$\ln k_a = \ln k_{a0} - \frac{U_0/k_B T}{1+\kappa a} \qquad (23)$$

where U_0 and a are fitting parameters. Comparison of eq 23 with eq 20 shows that U corresponds to the electrostatic interaction energy of the transient complex at $\kappa = 0$ (i.e., in the absence of salt). According to eq 23, a plot of ln k_a versus $(1 + \kappa a)^{-1}$ is linear. The intercept of the line at $(1 + \kappa a)^{-1} = 0$ is ln k_{a0}, from which the basal rate constant k_{a0} can be determined (Figure 7). The y value at $(1 + \kappa a)^{-1} = 1$ is ln k_a in the absence of salt, wherein the electrostatic attraction is maximized.[134] This linear relation was shown to hold for the association of the TEM1−BLIP, interferon−receptor, hirudin−thrombin, and barnase−barstar pairs and a heterodimeric leucine zipper for all salt concentrations tested.[93,129,135]

Figure 7 clearly shows that electrostatic attraction may be very strong at low salt. This phenomenon is well known for proteins and the basis of ion-exchange purification. To avoid strong, nonspecific electrostatic attraction the ionic strength in the cell is ∼150 mM. At this ionic strength charge−charge interactions are partially shielded, reducing the negative effect of nonspecific interactions by allowing for specific rate enhancement. A good example for this was reported for the complex of barnase−barstar in the presence of the poly ion hirudin.[136] The association rate at low salt was actually lower than in higher salt due to nonspecific interactions of barnase with hirudin, which effectively removed free barnase from the system. The rate peaked at 150−200 mM salt and slowed down at higher salt (this time due to masking of the charges). Thus, the physiological ionic strength is optimal to obtain fast specific binding yet reduce nonspecific binding.

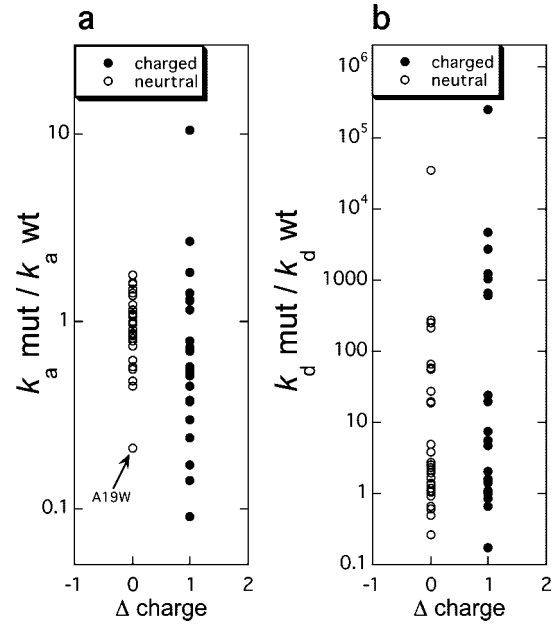


Figure 8. Changes in the rate constants of (a) association and (b) dissociation plotted against the change in charge resulting from a mutation. The rate constants were measured for 55 mutations in the barnase−barster, TEM1−BLIP, and IFNα2−IFNAR2 pairs.

4.2. Probing the Contributions of Individual Residues Toward k_a by Mutation

Measuring the effects of mutations on the rate constants of association is a powerful tool to decipher the mechanism of association. As already discussed in section 3.2, extensive site-directed mutagenesis of surface residues of the TEM1−BLIP, barnase−barstar, interferon−receptor, growth-hormone−receptor, IL4−receptor, and other protein pairs has demonstrated that mutations involving charged residues have the largest effect on k_a, while mutations of uncharged resides are much less important.[6,93,107,108,135,137−141] This point is further illustrated in Figure 8, which presents the analysis on the effects on k_a and k_d by a set of 55 single mutations located in the binding interfaces of the TEM1−BLIP, barnase−barstar, and interferon−receptor pairs. Except for the A19W mutation in interferon, none of the neutral mutations had a large effect on k_a, while many of them had major effects on k_d. This points toward a clear distinction between the process of association and dissociation. Hotspot residues affecting dissociation rates are both of charged and uncharged nature, while hotspots for association are almost always charged residues. Moreover, the energy gained or lost though a mutation is of a different scale, while for k_a it is rare to find a residue that alters the rate by over 10-fold; for k_d it is common that >5 such residues are found within an interface. While charge is important for association, the magnitude of perturbation of k_a is not simply a measure of the change in charge but rather relates to the specific location of the mutation and its contribution to the electrostatic energy of interaction between the two proteins in the transient complex (eq 20).[1,13,77,93,94] For example, charged mutations located far away from the binding interface will not affect association even if these mutations are drastic.[6]

4.2.1. Altering Association Kinetics by Protein Design

As will be discussed in section 4.3, eq 20 allows for a rigorous prediction of the association rate constant. The prediction requires generation of the transient complex. For

design purposes a simplified approach is called for. In the simplified approach taken by Schreiber and co-workers the electrostatic interaction energy $\langle U_{el} \rangle^*$ was calculated on the native complex itself[93,94]

$$\langle U_{el} \rangle^* \approx U_C - U_A - U_B \quad (24)$$

Here U_X, X = A, B, or C, refers to the electrostatic energy of two proteins or their native complex given by

$$U_X = \frac{1}{2} \sum_{i,j} \frac{q_i q_j e^{-\kappa(r_{ij}-a)}}{\varepsilon_s r_{ij} \, 1+\kappa a} \quad (25)$$

where q_i are the charges of the atoms in a protein or the protein complex and r_{ij} are the distances between the charges. The parameter a is the same as the one appearing in eq 23, and a value of 5.6 Å was found experimentally to give the best results in the calculations.[93,94] Despite their simplicity and the use of the native complex instead of the transient complex, eqs 24 and 25 were shown to give good estimates of electrostatic contributions to the rate constants of association and of the way mutations will affect the rate constants for diverse systems such as RNaseA−RI, AChE−fas, TEM1−BLIP, hiruding−thrombin, CheY−CheA, Ras−Raf, and Ras−Ral (Figure 5b).[9,14,93,94,132,142] This was found to be true whether mutations were placed within or outside the binding interfaces. Most informative were the cases where charged mutations were engineered outside the binding interfaces of protein−protein complexes, such as for the TEM1−BLIP and Ras−Ral pairs.[14,93] In these cases, strong increases in the rate constants of association were achieved (250- and 17-fold, respectively), in agreement with the calculated values. These increased k_a values were however not accompanied by a change in k_d, leading to increased association constants of the magnitudes described. This observation has far-reaching implications for our understanding of the transient complex and transition state for association, as will be described below. It is important to note that eqs 24 and 25 successfully predict the changes in association rate constant also for mutations located within the binding interfaces of protein−protein complexes, as shown for the association of the barnase−barstar, TEM1−BLIP, Ras−Ral, AChE−fas, hirudin−thrombin, CheY−CheA, and other protein pairs.[6,132]

On the basis of eqs 24 and 25, two web servers for calculating the changes in k_a upon mutations were set up: PARE (http://www.weizmann.ac.il/home/bcges/PARE.html) and HyPare (http://bip.weizmann.ac.il/HyPare). The first provides a more exact calculation for single mutations, while the second provides a full analysis of the entire protein surface, including the automated identification of "hotspots" for association.[94,143]

According to eq 20, basal rate constants are the rate constants of association in the absence of electrostatic attraction. These can be calculated either by extrapolating the values of k_a to the limit of infinite salt concentration or by introducing mutations that reduce the electrostatic interaction energy to zero. From such calculations the basal rate constants were found to be 4.4×10^4 or 4×10^4 M^{-1} s^{-1} for thrombin−hirudin association, 4.2×10^4 or 2.3×10^4 M^{-1} s^{-1} for TEM1−BLIP association, and 1.4×10^7 or 1.5×10^5 M^{-1} s^{-1} for barnase−barstar association (the two values are extrapolated from mutant or salt data, respectively, see Figure 7 and eq 23). Similarly, the values of k_{a0} were 6.6×10^5 M^{-1} s^{-1} for Ras−Raf association and 2×10^5 M^{-1} s^{-1} for AChE−fas association using mutation data. Except for barnase−barstar association, the basal rate constants were the same whether calculated from salt or mutant data. The reason for the large difference between the two methods in predicting the basal rate constant for barnase−barstar association is currently not clear;[143] possibly this is an indication of the limitation of the simplification in eqs 24 and 25 (see section 4.3 for further discussion).

It is important to note that eq 25 does not treat explicitly the contributions of noncharged residues to k_a. While indeed the contributions are small, they were found to be significant in a number of cases. For example, the A19W mutation on IFNα2 reduced k_a by 4-fold, a reduction that clearly relates to conformational rearrangement during the process of association (as verified using double-mutant cycle analysis with the W100A mutation on IFNAR2).[144] Another example for conformational rearrangement rather than electrostatics affecting association rates upon mutation was shown for MICA (a major histocompatibility complex-like protein) undergoing a structural transition from disorder to order upon binding its immuno-receptor, NKG2D.[145] In contrast, the rate-limiting step for the human transcription factor pKID domain of CREB, which also undergoes a folding transition upon binding to the KIC domain of the coactivator CBP, appears to be formation of a transient complex.[146,147] Stabilizing the structure of the pKID domain prior to binding does not increase the association rate constant. Also, for complex formation between fas and AChE, molecular dynamics simulations suggested that fas undergoes a conformational rearrangement during the binding to AChE, but the rearrangement does not appear to slow down the association rate.[148,149]

4.2.2. Proteomics View of 'Hotspots' for Association

As discussed above, the association rate constant of a protein complex depends on a basal rate constant and the magnitude of electrostatic attraction, which can be calculated in a simplified fashion from the structure of the native complex. In an analysis on a database of 68 heterodimeric complexes using HyPare it was found that in about one-half of the complexes electrostatic contributions to the association rate constants were small (<10-fold effects on k_a); in one-quarter of the complexes electrostatic attraction had a major effect on k_a (>100-fold increases). Defining a residue as a hotspot for association if it changes k_a by over 10-fold leaves about one-half of the complexes without any potential hotspot and a few hotspots per complex in the other complexes. Of these putative hotspot residues, about 40% are calculated to increase the rate of association upon mutation and thus increase binding affinity. These data suggest that a majority of protein−protein complexes are not optimized for fast association. Moreover, about 40% of the hotspots for association are located outside the binding interfaces, making them ideal candidates for protein engineering to achieve faster and tighter binding protein complexes. Hotspot residues are not evenly distributed among the 20 types of amino acids. About 75% of all hotspots are charged residues. This is understandable as a charge-reversal mutant changes the total charge by two. More intriguing is the small number of hydrophobic residues that are hotspots, in comparison to polar residues. For 18 out of the 68 complexes in the database experimental values of k_a were available, from which basal rate constants were calculated to be in the range from 10^4 to 10^7 M^{-1} s^{-1}. Some of these basal rate constants were corroborated by both salt and mutant data (see above). The

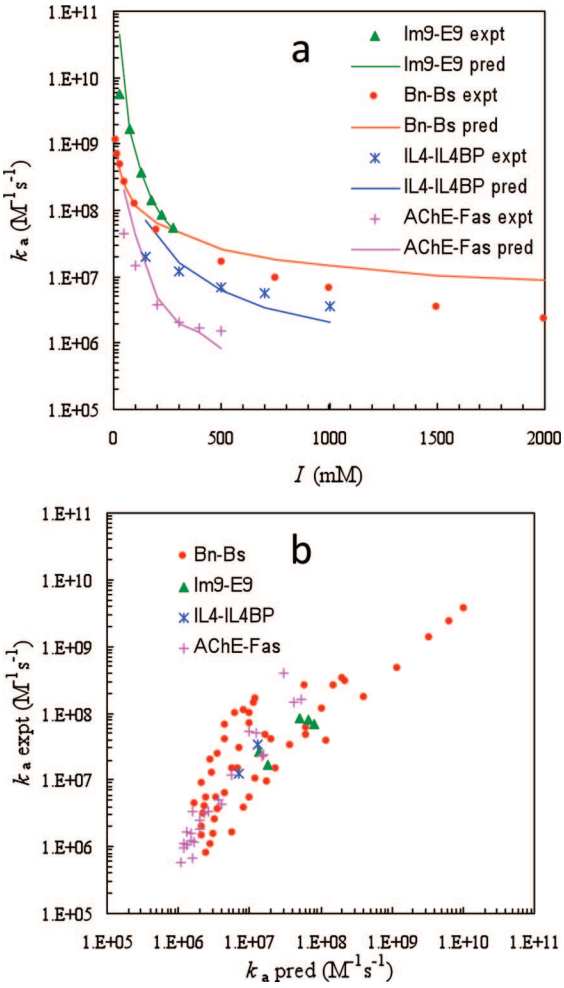


Figure 9. Comparison of predicted and experimental results for association rate constants. (a) Ionic-strength dependences of four protein pairs. (b) Twenty-three mutants at various ionic strengths. (Reprinted with permission from ref 13. Copyright 2007 Elsevier.)

basal rate constants seem to correlate with the sizes of the proteins and shapes of the interfaces (see the next section).

4.3. Prediction of Electrostatic Rate Enhancement by the Transient-Complex Theory

Let us now discuss the rigorous prediction of the association rate constant from the transient-complex theory. The theory is given by eq 20 along with a structural model for the ensemble of the transient complex, which is described in section 3.1. Transient-complex theory offers two important advantages. First, unlike in previous theoretical approaches the model for the transient complex is based on rigorous theoretical considerations and is uniquely defined. Second, electrostatic interactions between two associating proteins are treated directly by the Poisson−Boltzmann equation without approximations.

Transient-complex theory was put to a comprehensive test against experimental data[4−6,128] for the association rate constants of four protein pairs (shown in Figure 1) and 23 of their mutants over wide ranges of ionic strength.[13] For each protein pair 100 configurations were randomly selected from the transient-complex ensemble to calculate the average electrostatic interaction energy, $\langle U_{el}\rangle^*$. Equation 20 was then used to predict the association rate constant. As shown in Figure 9a the ionic strength dependences of the association rates for all four protein pairs are predicted well by the transient-complex theory. Moreover, the predictions for 23 mutants at various ionic strengths agree closely with experimental results (Figure 9b). In all there are 81 data points in the latter comparison, spanning 4 orders of magnitude in association rate. The theory thus appears to fulfill the promise of having truly predictive power. It reveals that among the protein pairs and their mutants studied the basal rate k_{a0} can differ by ∼20-fold, but the bulk of the variations in k_a is due to the variations in $\langle U_{el}\rangle^*$, which ranges from 0 to −6 kcal/mol (the last value translates into a 10^4-fold rate enhancement).

In the rigorous calculation of k_a it was found that the same basal rate constant, 1.4×10^6 M^{-1} s^{-1}, explains both salt and mutation data on barnase−barstar association. The relatively high basal rate constant for the barnase−barstar complex compared to other protein pairs is reproduced by Brownian dynamics simulations.[1,13] The variations of basal rate constants among protein complexes can be explained by the structures of the transient complexes. A higher k_{a0} value, as found for barnase−barstar association, corresponds to a more open ensemble of the transient complex.[1]

The question of whether the predicted k_a is sensitive to the precise specification of the transient complex was also addressed.[13] As explained in section 3.1, the transient complex is specified by the contact level, N_c^*, at the onset of a sharp increase in transltional/rotational freedom (Figure 3b). For the barnase−barstar complex the value of N_c^* was determined to be 14. The value of N_c^* was artificially varied from 10 to 18, and at each value k_{a0} and $\langle U_{el}\rangle^*$ were calculated. When combined according to eq 20, the predicted k_a was found to vary little (<2-fold) due the opposite dependences on N_c^* of the two contributing factors, k_{a0} and $\exp(-\langle U_{el}\rangle^*/k_BT)$, to k_a.

The comparison of theory against experimental data presented above was based on calculating the electrostatic interaction energy from the linearized Poisson−Boltzmann equation (eq 12). It has been found that when the full Poisson−Boltzmann equation (eq 11) was used, agreement with experiment improved, albeit modestly.[1] This underscores the point that a rigorous treatment of electrostatic interactions is essential for the accuracy of calculated k_a.

For the binding between a protein and an RNA the difference between the full Poisson−Boltzmann equation and the linearized version is no longer modest because of the large charge density on the nucleic acid. Then use of the full Poisson−Boltzmann equation becomes a necessity. Transient-complex theory has made it possible to realistically model protein−RNA binding rate constants for the first time.[26] In that work the binding of the spliceosomal protein U1A and its target on the U1 small nuclear RNA (Figure 10a) was studied. The binding and dissociation rate constants of this and other protein−RNA systems exhibit the disparate dependences on salt familiar to proteins,[150−153] indicating that the structural model for the transient complex developed for protein−protein association is applicable to protein−RNA binding. Representative configurations in the transient complex of the U1 system are shown in Figure 10b. As Figure 11 shows, the binding rate constants of the wild-type system and eight of its mutants are in close agreement with experimental data.[150,151]

Comparison of predicted and experimental association rates also helps settling an important technical detail in the calculation of the electrostatic interaction energy. The detail

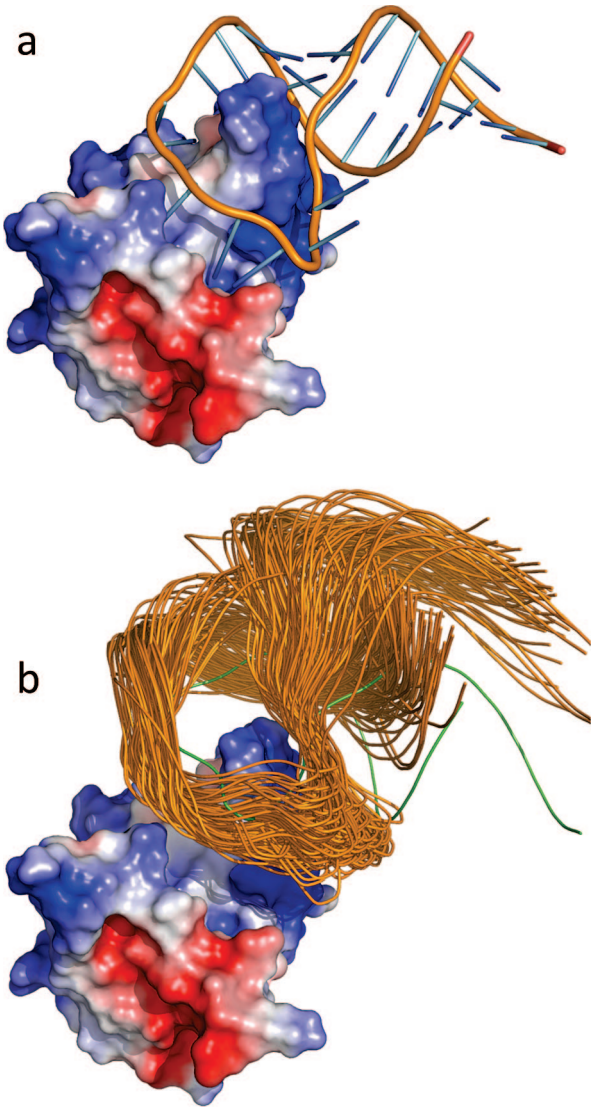


Figure 10. U1 protein−RNA system. (a) The native complex. (b) Representative configurations in the transient complex. The U1A protein is represented by the electrostatic surface. Notice that the RNA is moved away from the protein in the transient complex. (Reprinted with permission from ref 26. Copyright 2008 American Chemical Society.)

in question is the definition of the boundary between the protein low dielectric and solvent high dielectric. In many Poisson−Boltzmann calculations this dielectric boundary is specified by the molecular surface. An alternative choice is the van der Waals surface.[154,155] The difference between the two surfaces lies in the many crevices accessible to a spherical solvent probe, which are treated as part of the solute dielectric in the molecular surface protocol but as part of the solvent dielectric in the van der Waals surface protocol. It is known that the sign of the electrostatic interaction energy of two oppositely charged subunits can be reversed between the two protocols, positive in the former and negative in the latter.[110,156−158]

The results shown earlier for the average electrostatic interaction energies in the transient complexes were all obtained with the van der Waals surface protocol. With the molecular surface protocol the same sign reversal mentioned above was seen on $\langle U_{el}\rangle^*$.[1,26] Hence, electrostatic rate enhancement now turned into rate retardation. For example, for the barnase−barstar pair when the ionic strength was varied from 13 to 2000 mM $\langle U_{el}\rangle^*$ calculated with the van

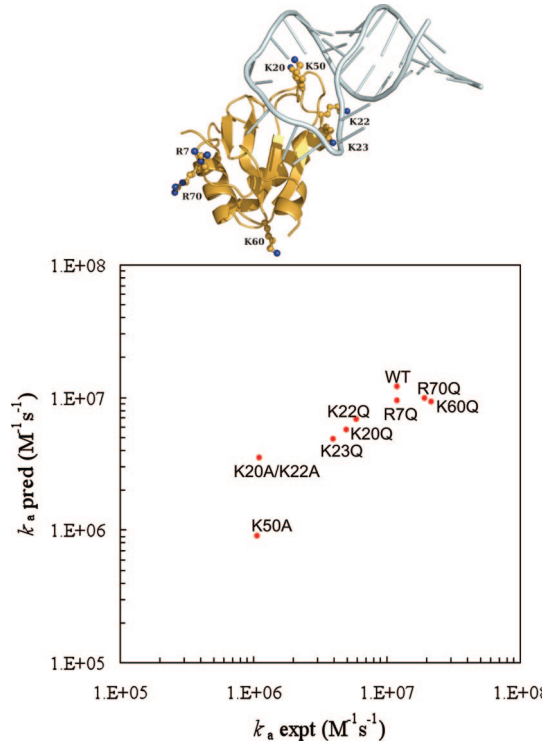


Figure 11. Comparison of predicted and experimental results for the binding rates of the wild-type U1 system and eight of its mutants. Locations of the mutated residues are shown in the inset. (Reprinted with permission from ref 26. Copyright 2008 American Chemical Society.)

der Waals surface protocol varied from −3.30 to −0.82 kcal/mol. Correspondingly, $\langle U_{el}\rangle^*$ calculated with the molecular surface protocol varied from 2.50 to 5.13 kcal/mol. For the latter results to be consistent with the experimental data for the association rate constant would require a basal rate constant in the order of $10^{10}-10^{11}$ $M^{-1}s^{-1}$, which is clearly unrealistic.

The transient-complex theory lays out a framework for understanding the wide variation in association rate constants (Figure 1) and elucidating the pathways of protein association and dissociation. As documented above, the theory rationalizes the insensitivity of k_d to ionic strength and has achieved great success in making quantitative predictions of k_a for association reactions where conformational rearrangements are not rate limiting. However, quantitative calculations for rate constants that involve formation or breakup of short-range stereospecific interactions are beyond the scope of the transient-complex theory.

5. Interaction Dynamics of Membrane-Anchored Proteins

Key processes in the initiation of transmembrane signaling involve ligand-induced interactions between receptor proteins, which are anchored to the membrane. Compared to interaction kinetics in solution, anchoring of proteins into the membrane dramatically changes the mobility of the interaction partners both in terms of diffusion constants (which are reduced by 2 orders of magnitude) and in restricting the movements to two dimensions (thus actually making the mutual search problem simpler). In general, reduction in dimensionality as well as potentially favorable preorientation has been proposed to explain the rapid initiation of signaling.[37,159−161] However, few quantitative

experimental studies of 2-dimensional protein interactions have been reported. For IL-4-induced cross-linking of the IL-4 Rα with the common γ chain on live cells a surprisingly low efficiency of association of the low-affinity γ-chain subunit with the highly stable complex of IL-4 and IL-4 Rα on the membrane was observed.[162] Recent systematic studies of ternary cytokine−receptor complexes tethered on artificial membranes have provided some quantitative understanding of 2-dimensional interactions. These data confirmed the regulatory role of binding affinities for complex formation on membranes. Equilibrium dissociation constants on the order of 1−10 μM for the interaction in solution yielded 2-dimensional equilibrium dissociation constants on the order of 10−100 molecules/μm^2.[163] Kinetic analysis of such complexes revealed that the probability of successful collisions is indeed higher on membranes, which can be ascribed to the preorientation and longer lifetime of the transient complex. This effect is, however, largely compensated by the decreased diffusion rate, which reduces the number of collisions on the same order of magnitude. The orientation of the proteins significantly affects the rate constants of association,[164] though protein domains interacting outside the membrane are often rather flexible. Overall, the association rates at a given average distance of the molecules in solution and on the surface are surprisingly similar.[165] In the dissociation process separation of the interacting species by diffusion plays a critical role. This agrees with the notion that the association kinetics is highly diffusion controlled. As a consequence, local membrane fluidity in membrane microdomains may play a critical role for stabilizing membrane protein complexes. Thus, the energy landscapes of cellular membrane protein complexes are highly influenced by the microcompartmentalization of the membranes. Crowding by by-stander proteins in the membranes may also have a significant effect (cf. section 6).

Single-molecule fluorescence techniques may contribute to the study of association within membranes. The process of dimerization of epidermal growth factor receptors (EGFR) in cellular membranes and binding of their substrate, epidermal growth factor (EGF), was studied recently using single-molecule fluorescence.[166] Fluorescent EGF molecules bound to EGFR molecules were tracked. Association of a second EGF molecule to a preformed EGF-EGFR complex was seen as a sudden doubling of the fluorescence intensity. It was concluded that EGFR dimerization preceded binding of a second EGF molecule to a complex. The kinetics of the dimerization process itself, however, is yet to be studied.

6. Roles of the Depletion Effect and Crowding in Association Kinetics

Since protein−protein association mostly takes place within cells it is important to understand the particular properties of the cellular environment which may have a profound effect on such reactions.[167] Most importantly, the cytoplasm of cells is a dense solution of macromolecules of many sizes and shapes. In fact, macromolecules may occupy as much as 40% of the total intracellular volume.[168] These "crowding agents", as they are sometimes called, include mostly folded proteins and nucleic acids. Structural proteins form dynamic networks of filaments, creating the cytoskeleton and thus dividing the cell into discrete bulk areas of cytoplasm, where soluble proteins can interact. In addition, there are many 'natively unfolded proteins' whose random-coil-like conformations can add complexity to the cytoplasmic milieu.[169]

There are two major outcomes of the high intracellular concentration of macromolecules in their various forms. First, the macroscopic viscosity of the cytoplasm is significantly higher than the viscosity of a dilute aqueous solution. This higher viscosity should affect long-range translational diffusion of proteins. On small length scales there should be a large heterogeneity in the density of macromolecules, leading to microscopic viscosity, which depends strongly on position. The crowded cytoplasm may also affect rotational motion of proteins, which is also crucial for protein−protein association. Since rotational diffusion is by nature more local than translational diffusion and occurs on faster time scales, it is likely to be affected differently by the crowded cytoplasm.[170,171] A second effect of the crowded cytoplasmic solution is depletion, which can be seen as an effective osmotic pressure exerted on all diffusing molecules. This phenomenon, which is entropic in nature, is due to the fact that protein solutes exclude (or deplete) some of the solution volume from other solutes. These effects are the subjects of the following sections.

6.1. Translational and Rotational Diffusion of Proteins in Concentrated Solutions and Cells

Dense macromolecular solutions should be discussed more appropriately in terms of their viscoelasticity rather than their viscosity.[172] Viscoeleastic solutions possess time-scale- and length-scale-dependent viscosity. The effective viscosity sensed by a tracer molecule undergoing fast, short-range motion is therefore different than that sensed by a particle moving slowly over large distances. Further, the cytoplasmic environment is not homogeneous, and the density of macromolecules and other intracellular components may vary from one spot to another within the cell. Thus, diffusion times may scale with distance in an unexpected way. Recent experiments have probed the viscoelasticity of cellular interiors using microrheology techniques (for reviews see refs 173 and 174). Of most interest to the present review are experiments that directly measure diffusion coefficients inside cells or in complex solutions that may mimic the cytoplasm.

A useful probe for studying translational and rotational diffusion inside cells is green fluorescent protein (GFP), which can be expressed alone or in tandem with other proteins. Verkman and co-workers used GFP and GFP-labeled proteins to study diffusion in the cytoplasm[170] and inside mitochondria.[175] They found that translational diffusion is more retarded than rotational diffusion, an observation which is in line with studies in model solutions described below. Particles larger than the ∼28 kDa GFP may show an even more complex behavior. Golding and Cox[176] studied a complex of ∼100 nm in size formed by a mRNA molecule with a large number of GFP-labeled proteins. The slow motion of this complex allowed the tracking of individual copies using time-lapse microphotography. The complex exhibited anomalous subdiffusive behavior with a mean-squared displacement proportional to t^{α}, where t is the time and $\alpha \approx 0.7$. This behavior suggests that the molecules are hindered in their motion, perhaps due to transient interactions with other components of the cell. Fluorescence correlation spectroscopy (FCS) of dextran molecules in cytoplasms also showed anomalous subdiffusion.[177]

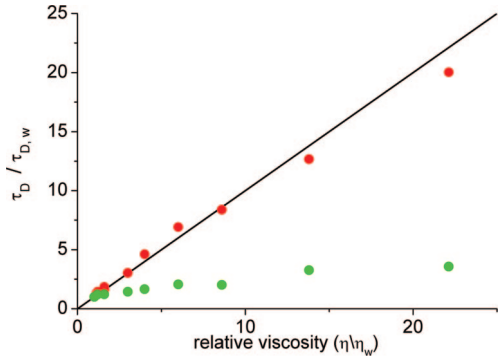


Figure 12. Protein translational and rotational correlation times in solutions of PEG of MW 8000, plotted as a function of solution viscosity. Translational correlation times (red circles) were measured by fluorescence correlation spectroscopy, while rotational correlation times (green circles) were measured by fluorescence polarization spectroscopy. All plotted values are normalized by water values. (Reprinted with permission from ref 180. Copyright 2008 American Chemical Society.)

To gain a broader view of diffusion in cell-like environments it is useful to conduct experiments in solutions of polymers in which the composition and concentration of cosolutes can be systematically changed.[178] Motion of proteins in solutions of large polymers is expected to be subdiffusive, and this behavior has been observed by Banks and Fradin,[179] who used FCS to study the motion of tracer proteins in solutions of dextran molecules of various sizes. Diffusion became more anomalous as the size of the polymer molecules and their concentration increased. This behavior was seen even in solutions of the smaller dextrans with molecular weights < 10 kDa. In contrast, in a study of diffusion of proteins in polyethylene glycol (PEG) solutions of molecular weights up to 8000 by FCS it was found that normal diffusion (i.e., mean-squared displacement linear with time) could well account for the data.[180] It is likely that the size of the crowding agents relative to the tracer protein size is the key parameter that determines the functional dependence of diffusive motion on time.

The relation of diffusion coefficients to viscosity can be another useful indicator of the degree of retardation of diffusive motion by polymeric solutions. The translational diffusion coefficients of small proteins in PEG solutions of molecular weights up to 8000 obeyed the Stokes−Einstein relation, i.e., in inverse proportion to solution macroscopic viscosity.[180] In stark contrast, rotational diffusion correlation times in the same solution were not proportional to macroscopic viscosity,[180] in agreement with other experiments.[181] The results for diffusion in PEG8000 are shown in Figure 12. Rotational motion occurs on a much faster time scale than translational motion, and therefore, rotation probes only the local viscosity at the position of a protein molecule, which experiences rather little friction from the polymer molecules themselves. This effect of microviscosity on protein rotational motion has a strong impact on protein−protein association in polymer solutions, as will be seen below.

6.2. Depletion Effect

The interaction between solutes in a concentrated solution can be understood within the framework of the classical McMillan−Mayer theory of multicomponent systems.[182] In the simplest form of this theory solute molecules are represented as hard spheres that cannot penetrate each other, but there is no direct interaction between them when they

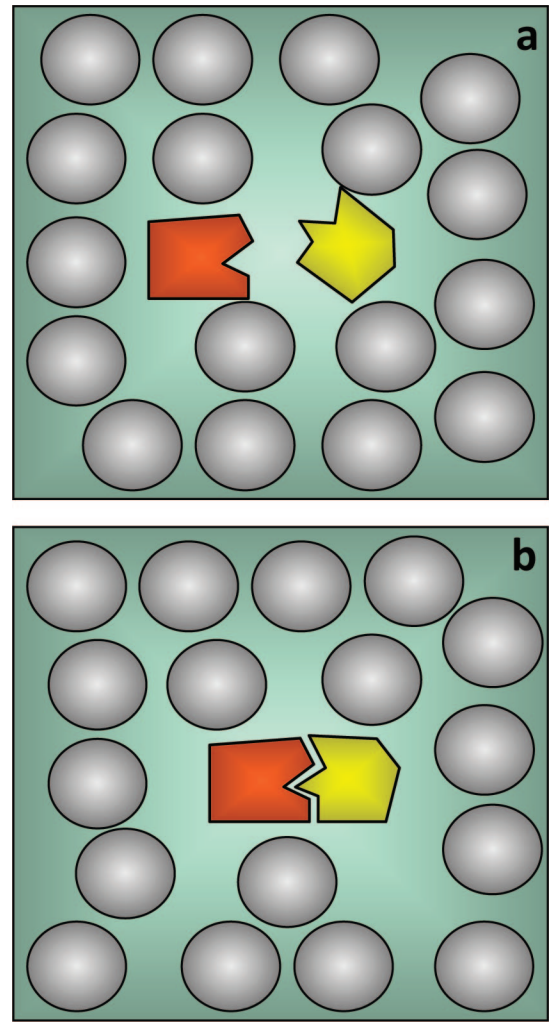


Figure 13. Depletion effect. (a) An osmotic pressure is exerted on two particles (proteins) immersed in a macromolecular solution when their separation does not allow the solute macromolecules to penetrate the volume between them. (b) This entropic effect is minimal when the two particles are in contact (e.g., when two proteins associate).

are separated. Their interactions and reaction equilibria, as they are modified due to the excluded volume effects of other solutes, can then be calculated.[183] Asakura and Oosawa realized, in a terse letter to the *Journal of Chemical Physics* written in 1954,[184] that the excluded volume effect on two particles immersed in a macromolecular solution can be cast in terms of an effective distance-dependent attractive potential, which is operating when their distance is smaller than the size of a macromolecule. This attractive potential[185] can be readily attributed to the inability of the macromolecular solute molecules to enter the region between the two particles (Figure 13). This region is said to be 'depleted' of solute molecules; hence, the name 'depletion effect' given to this phenomenon. The depletion potential, entropic in nature, will prove useful for calculating the effect of polymer cosolutes on protein−protein association kinetics (section 6.3).

The depletion effect is similar to the situation occurring in an osmotic pressure experiment, where a semipermeable membrane precludes macromolecular solutes from entering one region of the solution. In that Asakura and Oosawa (AO) theory the depletion potential is indeed proportional to the osmotic pressure of the solution as well as to the depleted volume. The AO theory is formally correct only in the colloid

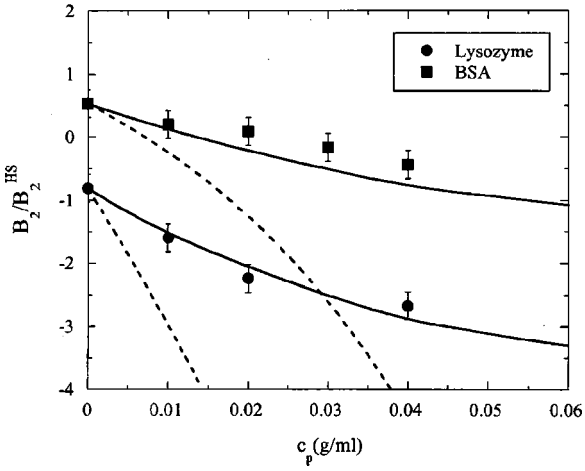


Figure 14. B_2 (normalized by the hard-sphere value) for lysozyme and BSA as a function of concentration (wt %) of PEG of molecular weight 12 000. The dotted and solid lines are AO and PRISM predictions, respectively. (Adapted with permission from ref 193. Copyright 2000 American Institute of Physics.)

limit, when the macromolecules exerting the depletion force are much smaller than the particles feeling it.[186] It also does not take into account the polymeric nature of macromolecular solutions, modeling the macromolecules as hard spheres. Schweizer and co-workers[187,188] developed an integral equation theory, called the Polymer Reference Interaction Site Model (PRISM), which attempts to properly treat macromolecular solutions and is also applicable beyond the colloid limit.

The excluded-volume interactions underlying the depletion potential manifest themselves in many equilibrium processes.[183] Indeed, precipitation and crystallization of proteins by polymeric additives have been studied for many years.[189] Minton and co-workers developed a sedimentation equilibrium method[190] to study excluded volume-driven association between proteins. In particular, inert proteins, acting as crowders, were shown to promote specific self-association of the cell division protein FtsZ into rod-like oligomers.[191]

Theoretically, equilibrium effects of excluded-volume interactions on a solution of proteins can be quantified through the second virial coefficient, B_2, which is related to the depletion potential, $U(r)$, via

$$B_2 = 2\pi \int_0^\infty r^2(1 - \exp[U(r)/k_BT]\mathrm{d}r \qquad (26)$$

The effect of crowding and depletion on the second virial coefficient of protein solutions was directly probed by scattering experiments.[192−195] Kulkarni et al. performed a series of experiments on the proteins lysozyme and bovine serum albumin in PEG solutions of several molecular weights.[193] The dependence of B_2 on polymer concentration was extracted and compared to models in order to understand nonspecific, solute-induced interactions between protein molecules. Since PEG is a flexible polymer, the AO theory of depletion did not describe the results well. It was found that the PRISM theory, which takes into account the flexibility of polymer, accounts better for the experimental results (Figure 14). Vivares et al. extracted the full $U(r)$ from small-angle X-ray scattering experiments of urate oxidase in PEG solutions.[195] By comparison to several models they also reached the conclusion that PEG-induced depletion has to be described by models that take into account its polymeric nature.

6.3. Crowding and Protein−Protein Association Kinetics

The kinetics of protein−protein association in a crowded environment may be affected by both effects discussed above.[196] First, the increase in solution viscosity should slow down diffusion, which is what brings two associating proteins toward their transient complex. Since translational and rotational diffusion may be differently affected by the presence of macromolecules (as discussed in section 6.2), modeling the viscosity effect on association rates may not be trivial. Second, the effective attraction induced by the depletion interaction may affect not only the equilibrium of protein solutions but also the kinetics of their interactions. Just as an electrostatic potential enhances the association rate (see sections 1.1 and 4), so is the depletion potential expected to speed up association. When two protein molecules are close enough to each other, the osmotic pressure exerted by macromolecular cosolutes should lead to a significant increase in the probability of reaching the transient complex. A simple albeit approximate description of the effect of depletion on association kinetics can be obtained using eq 17, where now $\langle U \rangle^*$ is interpreted as the average depletion interaction potential in the transient complex in which the two proteins are essentially at contact. Illustrative calculations in ref 196 show that when the effect of slowed down diffusion and the contribution of the attractive depletion potential are combined "cancellation of the two opposing effects leaves a modest overall effect of crowding on the binding rate".

Several studies of protein−protein association kinetics in macromolecular solutions have been published.[197−199] The common denominator of all these experiments is that the effect of polymers on association kinetics is small, which, as just noted, is precisely what is predicted by theory.[196] The first experimental evidence for the underlying reason in that prediction was obtained in a recent study in which the association rate was dissected into its various components.[180] This study is discussed next.

6.3.1. Case of TEM−BLIP

The association kinetics of the protein TEM1−β-lactamase (TEM1) and its protein inhibitor BLIP was studied in macromolecular solution in an attempt to mimic the crowded intracellular milieu and systematically probe its effects on the process. A stopped-flow apparatus was used to measure the millisecond kinetics of association of these proteins in solutions of polyethylene glycol molecules of various molecular weights from 200 to 8000 (designated as PEG200, PEG8000, etc.).[198] Measuring the translational and rotational diffusion coefficients in some of these solutions[180] allowed the use of eq 8 to provide an estimate for the change in the association rate constant due to changes in the diffusion coefficients. The reader is reminded that in this equation the association rate constant depends on both translational and rotational diffusion coefficients. If the diffusion coefficients both obey the Stokes−Einstein relations, i.e., are inversely proportional to solution viscosity, then eq 8 predicts that the association rates will also be inversely proportional to viscosity. A much weaker dependence on viscosity was seen in the experiments.[180,198]

It is useful to describe here the PEG8000 results, which were particularly clear and striking. When plotted as a function of viscosity the association times (inverse rates) in

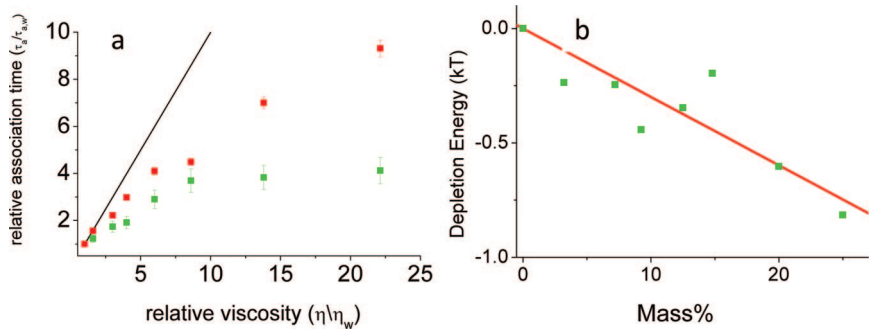


Figure 15. Effect of PEG8000 on association times of the TEM1−BLIP protein pair. (a) Measured association times (relative to water, green squares) and association times calculated from translational and rotational diffusion coefficients (red squares) as a function of relative viscosity. The black line is the Stokes−Einstein prediction. (b) Depletion energies obtained from the difference between measured and calculated association times are plotted as a function of polymer concentration. (Reprinted with permission from ref 180. Copyright 2008 American Chemical Society.)

PEG8000 solutions are much faster than the SE prediction and in fact seem to saturate and remain essentially constant above a relative viscosity value of ~5 (Figure 15A, green squares). This result seems to be in accord with the measurement of rotational diffusion times of the proteins in PEG8000 solutions, which also shows a saturating behavior (Figure 12). However, translational diffusion times obey the SE relation well, and when the various diffusion times are inserted into eq 8 the predicted association times are found to be too large (Figure 15A, red squares). The significant enhancement of association in PEG8000 solutions can be attributed to depletion, i.e., to the effective attraction induced by the polymer solution on the two proteins. Equation 17 can be used to estimate the interaction potential due to the depletion effect as a function of polymer concentration. This results in the depletion-induced interaction potential being linear with concentration (Figure 15B) in line with the AO theory. In other words, the depletion effect on association time is exponential in the polymer concentration. However, calculations using eq 17 show that the effect is significantly weaker than predicted by the AO theory, suggesting that the geometric considerations used to derive the depletion potential in that theory might be oversimplified. This is not surprising since PEG8000 molecules cannot be treated as spheres, certainly not at the concentrations used in the experiment. On the other hand, these results cannot be explained by the PRISM theory of Schweitzer and coworkers[187,188] either, which predicts a linear dependence of the depletion effect on concentration rather than an exponential dependence. Further theoretical analysis is thus required in order to fully rationalize the experimental findings.

The biological implications of the above results are quite interesting. Over a rather broad range of polymer concentrations the rate of association between proteins is likely to be only mildly affected by the large abundance of other macromolecules. This is due to the 'corrective' effect of the depletion interaction, which enhances association much beyond the diffusion-limited rate.[196]

6.3.2. Surprises in Concentrated Solutions

When the concentration of macromolecules in solution is so large that the polymer solution correlation length becomes comparable to the size of a monomer an interesting regime is approached. In this regime monomers from one chain interact with monomers on other chains in the same manner as they interact with themselves. This leads to an effective loss of the excluded volume interaction on internal statistics of polymer molecules, and the statistics become those of ideal chains.[200] The effect of concentrated macromolecular solutions on association reactions of solute proteins should be similar to the effect in concentrated solutions of the relevant monomers.

Studying the effect of concentrated macromolecular solutions on the kinetics of the TEM1−BLIP association reaction turned out to be problematic for larger polymers, e.g., PEG8000, due to the excessively high viscosity. Nevertheless, it was possible to study association in concentrated solutions of smaller polymers, e.g., PEG1000.[201] It was found that above a certain critical concentration of the polymer a dramatic increase in association times occurred. Quite satisfactorily, the dependence of the change in association times in this phase on polymer concentration was rather similar to the dependence of association times on concentration in solutions of the monomeric ethylene glycol.[201] This observation matches the physics of concentrated polymer solutions as described above. It is possible that in cellular environments a continuous change of the crowding effect, induced by changes in concentrations of proteins and other macromolecules, may strongly modulate association kinetics.

7. Conclusions

The association of proteins to form a complex is a multistep process, which starts by random collisions of the individual proteins. Multiple collisions and rotational diffusion brings the proteins to an orientation that is close to that of the native complex, leading to formation of a transient complex. This part of the process is diffusion controlled and strongly affected by electrostatic interactions. An open question is whether the transient complex develops into the native complex through a transition state or complex formation is a simple downhill transformation of the transient complex. In section 1.2 theoretical models are used to describe association as a diffusion-limited process, which is slowed down from the Smoluchowski limit due to orientational constraints satisfied by a stereospecific native complex. On the other hand, in section 3.2 experimental evidence is presented to show that the transient complex will likely dissociate and that the transition state for association is related to short-range conformational rearrangement and desolvation, processes that slow the conversion from the transient complex to the native complex. However, structurally the transient complex and transition state seem to be similar.

8. Acknowledgments

This work was supported in part by the Isreal Academy of Sciences and Humanities grant 389/02-1 to G.S. and NIH Grant GM58187 to H.-X. Z.

9. References

(1) Alsallaq, R.; Zhou, H.-X. *Proteins* **2008**, *71*, 320.
(2) Zhou, H.-X. *Phys. Biol.* **2005**, *2*, R1.
(3) Terlau, H.; Shon, K.-J.; Grilley, M.; Stocker, M.; Stuhmer, W.; Baldomero, O. M. *Nature* **1996**, *381*, 148.
(4) Radic, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. *J. Biol. Chem.* **1997**, *272*, 23265.
(5) Wallis, R.; Moore, G. K.; James, R.; Kleanthous, C. *Biochemistry* **1995**, *34*, 13743.
(6) Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.* **1996**, *3*, 427.
(7) Schreiber, G.; Buckle, A. M.; Fersht, A. R. *Structure* **1994**, *2*, 945.
(8) Shapiro, R.; Ruiz-Gutierrez, M.; Chen, C.-Z. *J. Mol. Biol.* **2000**, *302*, 497.
(9) Johnson, R. J.; McCoy, J. G.; Bingman, C. A.; Phillips, G. N., Jr.; Raines, R. T. *J. Mol. Biol.* **2007**, *368*, 434.
(10) Kim, A. S.; Kakalis, L. T.; Abdul-Manan, N.; Liu, G. A.; Rosen, M. K. *Nature* **2000**, *404*, 151.
(11) Hemsath, L.; Dvorsky, R.; Fiegen, D.; Carlier, M. F.; Ahmadian, M. R. *Mol. Cell* **2005**, *20*, 313.
(12) Marchand, J.-B.; Kaiser, D. A.; Pollard, T. D.; Higgs, H. N. *Nat. Cell Biol.* **2001**, *3*, 76.
(13) Alsallaq, R.; Zhou, H.-X. *Structure* **2007**, *15*, 215.
(14) Kiel, C.; Selzer, T.; Shaul, Y.; Schreiber, G.; Herrmann, C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9223.
(15) Korennykh, A. V.; Piccirilli, J. A.; Correll, C. C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 436.
(16) Park, C.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 11472.
(17) Perrett, S.; Zahn, R.; Stenberg, G.; Fersht, A. R. *J. Mol. Biol.* **1997**, *269*, 892.
(18) Alsallaq, R.; Zhou, H.-X. *Biophys. J.* **2007**, *92*, 1486.
(19) Berg, O. G.; von Hippel, P. H. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131.
(20) Zhou, H.-X.; Wong, K. Y.; Vijayakumar, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12372.
(21) Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. *Biophys. J.* **1999**, *76*, 1166.
(22) Zhou, H.-X. *Biopolymers* **2001**, *59*, 427.
(23) Huang, X.; Dong, F.; Zhou, H.-X. *J. Am. Chem. Soc.* **2005**, *127*, 6836.
(24) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917.
(25) Eyring, H. *J. Chem. Phys.* **1932**, *3*, 107.
(26) Qin, S. B.; Zhou, H.-X. *J. Phys. Chem. B* **2008**, *112*, 5955.
(27) Kramers, H. A. *Physica* **1940**, *7*, 284.
(28) Northrup, S. H.; Erickson, H. P. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3338.
(29) Zhou, H.-X. *Biophys. J.* **1997**, *73*, 2441.
(30) Camacho, C. J.; Kimura, S. R.; DeLisi, C.; Vajda, S. *Biophys. J.* **2000**, *78*, 1094.
(31) Schlosshauer, M.; Baker, D. *Protein Sci.* **2004**, *13*, 1660.
(32) Foote, J.; Eisen, H. N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1254.
(33) Hoffman, T. L.; LaBranche, C. C.; Zhang, W.; Canziani, G.; Robinson, J.; Chaiken, I.; Hoxie, J. A.; Doms, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6359.
(34) Wassaf, D.; Kuang, G.; Kopacz, K.; Wu, Q. L.; Nguyen, Q.; Toews, M.; Cosic, J.; Jacques, J.; Wiltshire, S.; Lambert, J.; Pazmany, C. C.; Hogan, S.; Ladner, R. C.; Nixon, A. E.; Sexton, D. J. *Anal. Biochem.* **2006**, *351*, 241.
(35) Darling, R. J.; Kuchibhotla, U.; Glaesner, W.; Micanovic, R.; Witcher, D. R.; Beals, J. M. *Biochemistry* **2002**, *41*, 14524.
(36) Uter, N. T.; Gruic-Sovulj, I.; Perona, J. J. *J. Biol. Chem.* **2005**, *280*, 23966.
(37) Zhou, H.-X.; Szabo, A. *Phys. Rev. Lett.* **2004**, *93*, 178101.
(38) Zhou, H.-X. *Biophys. J.* **2005**, *88*, 1608.
(39) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nucleic Acids Res.* **2001**, *29*, 2860–2874.
(40) Tjong, H.; Zhou, H.-X. *Nucleic Acids Res.* **2007**, *35*, 1465.
(41) Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69.
(42) Zhou, H.-X.; Wlodek, S. T.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9280.
(43) Abdul-Manan, N.; Aghazadeh, B.; Liu, G. A.; Majumdar, A.; Ouerfelli, O.; Siminovitch, K. A.; Rosen, M. K. *Nature* **1999**, *399*, 379.
(44) Hage, T.; Sebald, W.; Reinemer, P. *Cell* **1999**, *97*, 271.
(45) Smoluchowski, M. V. *Z. Phys. Chem.* **1917**, *92*, 129.
(46) Debye, P. *Trans. Electrochem. Soc.* **1942**, *82*, 265.
(47) Solc, K.; Stockmayer, W. H *Int. J. Chem. Kinet.* **1973**, *5*, 733.
(48) Berg, O. G. *Biophys. J.* **1985**, *47*, 1.
(49) Zhou, H.-X. *Biophys. J.* **1993**, *64*, 1711.
(50) Shoup, D.; Lipari, G.; Szabo, A. *Biophys. J.* **1981**, *36*, 697.
(51) Temkin, S. I.; Yakobson, B. I. *J. Phys. Chem.* **1984**, *88*, 2679.
(52) Schlosshauer, M.; Baker, D. *J. Phys. Chem. B* **2002**, *106*, 12079.
(53) Dafforn, T. R.; Smith, C. J. *EMBO Rep.* **2004**, *5*, 1046.
(54) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35.
(55) Jemth, P.; Gianni, S. *Biochemistry* **2007**, *46*, 8701.
(56) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121.
(57) Piehler, J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 4.
(58) Nieba, L. *Anal. Biochem.* **1996**, *234*, 155.
(59) Piehler, J.; Schreiber, G. *J. Mol. Biol.* **1999**, *289*, 57.
(60) Albeck, S.; Unger, R.; Schreiber, G. *J. Mol. Biol.* **2000**, *298*, 503.
(61) Rich, R. L.; Cannon, M. J.; Jenkins, J.; Pandian, P.; Sundaram, S.; Magyar, R.; Brockman, J.; Lambert, J.; Myszka, D. G. *Anal. Biochem.* **2008**, *373*, 112.
(62) Northrup, S. H.; Reynolds, J. C. L.; Miller, C. M.; Forrest, K. J.; Boles, J. O. *J. Am. Chem. Soc.* **1986**, *108*, 8162.
(63) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149.
(64) Altobelli, G.; Subramaniam, S. *Biophys. J.* **2000**, *79*, 2954.
(65) Fogolari, F.; Ugolini, R.; Molinari, H.; Viglino, P.; Esposito, G. *Eur. J. Biochem.* **2000**, *267*, 4861.
(66) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139.
(67) Gabdoulline, R. R.; Kummer, U.; Olsen, L. F.; Wade, R. C. *Biophys. J.* **2003**, *85*, 1421.
(68) Zou, G.; Skeel, R. D. *Biophys. J.* **2003**, *85*, 2147.
(69) Lin, J.; Beratan, D. N. *J. Phys. Chem. B* **2005**, *109*, 7529.
(70) Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Wade, R. C.; Helms, V. *Biophys. J.* **2006**, *90*, 1913.
(71) Haddadian, E. J.; Gross, E. L. *Biophys. J.* **2006**, *91*, 2589.
(72) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868.
(73) Northrup, S. H.; Allison, S. A.; McCammon, J. A. *J. Chem. Phys.* **1984**, *80*, 1517.
(74) Zhou, H.-X. *J. Chem. Phys.* **1990**, *92*, 3092.
(75) Zhou, H.-X. *J. Phys. Chem.* **1990**, *94*, 8794.
(76) Zhou, H.-X.; Szabo, A. *Biophys. J.* **1996**, *71*, 2440.
(77) Vijayakumar, M.; Wong, K. Y.; Schreiber, G.; Fersht, A. R.; Szabo, A.; Zhou, H.-X. *J. Mol. Biol.* **1998**, *278*, 1015.
(78) Zhou, H.-X. *J. Chem. Phys.* **1996**, *105*, 7235.
(79) Zhou, H.-X.; Briggs, J. M.; McCammon, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 13069.
(80) Zhou, H.-X.; Briggs, J. M.; Tara, S.; McCammon, J. A. *Biopolymers* **1998**, *45*, 355.
(81) Janin, J. *Proteins* **1997**, *28*, 153.
(82) Krishna, M. M.; Hoang, L.; Lin, Y.; Englander, S. W. *Methods* **2004**, *34*, 51.
(83) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
(84) Nolting, B.; Golbik, R.; Neira, J. L.; Soler, G. A.; Schreiber, G.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 826.
(85) Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 327.
(86) Haran, G. *J. Phys.: Condens. Matter* **2003**, *15*, R1291.
(87) Cornish, P. V.; Ha, T. *ACS Chem. Biol.* **2007**, *2*, 53.
(88) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123.
(89) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.
(90) Petrovich, M.; Jonsson, A. L.; Ferguson, N.; Daggett, V.; Fersht, A. R. *J. Mol. Biol.* **2006**, *360*, 865.
(91) Huang, X.; Zhou, H.-X. *Biophys. J.* **2006**, *91*, 2451.
(92) Miyashita, O.; Onuchic, J. N.; Okamura, M. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16174.
(93) Selzer, T.; Albeck, S.; Schreiber, G. *Nat. Struct. Biol.* **2000**, *7*, 537.
(94) Selzer, T.; Schreiber, G. *Proteins* **2001**, *45*, 190.
(95) Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33.
(96) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Recogit.* **1999**, *12*, 226.
(97) Volkov, A.; Worrall, J.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945.
(98) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383.
(99) Suh, J. Y.; Tang, C.; Clore, G. M. *J. Am. Chem. Soc.* **2007**, *129*, 12954.
(100) Keeble, A. H.; Kleanthous, C. *J. Mol. Biol.* **2005**, *352*, 656.
(101) Estrada, S.; Pavlova, A.; Björk, I. *Biochemistry* **1999**, *38*, 7339.
(102) Estrada, S.; Olson, S. T.; Raub-Segall, E.; Bjork, I. *Protein Sci.* **2000**, *9*, 2218.
(103) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.

(104) Mateu, M. G.; Sanchez Del Pino, M. M.; Fersht, A. R. *Nat. Struct. Biol.* **1999**, *6*, 191.
(105) Taylor, M. G.; Rajpal, A.; Kirsch, J. F. *Protein Sci.* **1998**, *7*, 1857.
(106) Wu, L. C.; Tuot, D. S.; Lyons, D. S.; Garcia, K. C.; Davis, M. M. *Nature* **2002**, *418*, 552.
(107) Reichmann, D.; Cohen, M.; Abramovich, R.; Dym, O.; Lim, D.; Strynadka, N. C.; Schreiber, G. *J. Mol. Biol.* **2007**, *365*, 663.
(108) Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995**, *248*, 478.
(109) Schreiber, G.; Fersht, A. R. *Biochemistry* **1993**, *32*, 5145.
(110) Dong, F.; Vijayakumar, M.; Zhou, H.-X. *Biophys. J.* **2003**, *85*, 49.
(111) Mejean, A.; Bodenreider, C.; Schuerer, K.; Goldberg, M. E. *Biochemistry* **2001**, *40*, 8169.
(112) Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835.
(113) Horovitz, A. *J. Mol. Biol.* **1987**, *196*, 733.
(114) Horovitz, A.; Fersht, A. R. *J. Mol. Biol.* **1990**, *214*, 613.
(115) Horovitz, A. *Fold Des.* **1996**, *1*, R121.
(116) Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. *Biochemistry* **1990**, *29*, 9343.
(117) Tetreault, M.; Cusanovich, M.; Meyer, T.; Axelrod, H.; Okamura, M. Y. *Biochemistry* **2002**, *41*, 5807.
(118) Harel, M.; Cohen, M.; Schreiber, G. *J. Mol. Biol.* **2007**, *371*, 180.
(119) Frisch, C.; Fersht, A. R.; Schreiber, G. *J. Mol. Biol.* **2001**, *308*, 69.
(120) Miyashita, O.; Onuchic, J. N.; Okamura, M. Y. *Biochemistry* **2003**, *42*, 11651.
(121) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(122) Zhou, H.-X. *Protein Sci.* **2003**, *12*, 2379.
(123) Miller, C. *Biochemistry* **1990**, *29*, 5320.
(124) Candia, S.; Garcia, M. L.; Latorre, R. *Biophys. J.* **1992**, *63*, 583.
(125) Escobar, L.; Root, M. J.; MacKinnon, R. *Biochemistry* **1993**, *32*, 6982.
(126) Goldstein, S. A. N.; Miller, C. *Biophys. J.* **1993**, *65*, 1613.
(127) Murrell-Lagnado, R. D.; Aldrich, R. W. *J. Gen. Physiol.* **1993**, *102*, 977.
(128) Shen, B. J.; Hage, T.; Sebald, W. *Eur. J. Biochem.* **1996**, *240*, 252.
(129) Wendt, H.; Leder, L.; Harma, H.; Jelesarov, I.; Baici, A.; Bosshard, H. R. *Biochemistry* **1997**, *36*, 204.
(130) Baerga-Ortiz, A.; Rezaie, A. R.; Komives, E. A. *J. Mol. Biol.* **2000**, *296*, 651.
(131) Walker, D.; Moore, G. R.; James, R.; Kleanthous, C. *Biochemistry* **2003**, *42*, 4161.
(132) Stewart, R. C.; Van Bruggen, R. *J. Mol. Biol.* **2004**, *336*, 287.
(133) Gianni, S.; Engstrom, A.; Larsson, M.; Calosci, N.; Malatesta, F.; Eklund, L.; Ngang, C. C.; Travaglini-Allocatelli, C.; Jemth, P. *J. Biol. Chem.* **2005**, *280*, 34805.
(134) Selzer, T.; Schreiber, G. *J. Mol. Biol.* **1999**, *287*, 409.
(135) Piehler, J.; Schreiber, G. *J. Mol. Biol.* **1999**, *294*, 223.
(136) Schreiber, G.; Frisch, C.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 111.
(137) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383.
(138) Horn, J. R.; Kraybill, B.; Petro, E. J.; Coales, S. J.; Morrow, J. A.; Hamuro, Y.; Kossiakoff, A. A. *Biochemistry* **2006**, *45*, 8488.
(139) Kouadio, J. L.; Horn, J. R.; Pal, G.; Kossiakoff, A. A. *J. Biol. Chem.* **2005**, *280*, 25524.
(140) Kraich, M.; Klein, M.; Patino, E.; Harrer, H.; Nickel, J.; Sebald, W.; Mueller, T. D. *BMC Biol.* **2006**, *4*, 13.
(141) Piehler, J.; Roisman, L. C.; Schreiber, G. *J. Biol. Chem.* **2000**, *275*, 40425.
(142) Marvin, J. S.; Lowman, H. B. *Biochemistry* **2003**, *42*, 7077.
(143) Shaul, Y.; Schreiber, G. *Proteins* **2005**, *60*, 341.
(144) Slutzki, M.; Jaitin, D. A.; Yehezkel, T. B.; Schreiber, G. *J. Mol. Biol.* **2006**, *360*, 1019.
(145) Lengyel, C. S.; Willis, L. J.; Mann, P.; Baker, D.; Kortemme, T.; Strong, R. K.; McFarland, B. J. *J. Biol. Chem.* **2007**, *282*, 30658.
(146) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021.
(147) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comput. Biol.* **2008**, *4*, e1000060.
(148) Bui, J. M.; Radic, Z.; Taylor, P.; McCammon, J. A. *Biophys. J.* **2006**, *90*, 3280.
(149) Bui, J. M.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15451.
(150) Law, M. J.; Linde, M. E.; Chambers, E. J.; Oubridge, C.; Katsamba, P. S.; Nilsson, L.; Haworth, I. S.; Laird-Offringa, I. A. *Nucleic Acids Res.* **2006**, *34*, 275.
(151) Katsamba, P. S.; Myszka, D. G.; Laird-Offringa, I. A. *J. Biol. Chem.* **2001**, *276*, 21476.
(152) Milev, S.; Bosshard, H. R.; Jelesarov, I. *Biochemistry* **2005**, *44*, 285.
(153) Auweter, S. D.; Fasan, R.; Reymond, L.; Underwood, J. G.; Black, D. L.; Pitsch, S.; Allain, F. H. *EMBO. J.* **2006**, *25*, 163.
(154) Vijayakumar, M.; Zhou, H.-X. *J. Phys. Chem. B* **2001**, *105*, 7334.
(155) Dong, F.; Zhou, H.-X. *Biophys. J.* **2002**, *83*, 1341.
(156) Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **2004**, *87*, 1618.
(157) Dong, F.; Zhou, H.-X. *Proteins* **2006**, *65*, 87.
(158) Qin, S. B.; Zhou, H.-X. *Biopolymers* **2007**, *86*, 112.
(159) Adam, G.; Delbruck, M. In *Structural Chemistry and Molecular Biology*; Rich, A., Davidson, N., Eds.; Freeman and Company: San Francisco, 1968.
(160) Axelrod, D.; Wang, M. D. *Biophys. J.* **1994**, *66*, 588.
(161) Wiegel, F. W.; DeLisi, C. *Am. J. Physiol.* **1982**, *243*, R475.
(162) Whitty, A.; Raskin, N.; Olson, D. L.; Borysenko, C. W.; Ambrose, C. M.; Benjamin, C. D.; Burkly, L. C. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13165.
(163) Gavutis, M.; Lata, S.; Lamken, P.; Muller, P.; Piehler, J. *Biophys. J.* **2005**, *88*, 4289.
(164) Lamken, P.; Gavutis, M.; Peters, I.; Van der Heyden, J.; Uze, G.; Piehler, J. *J. Mol. Biol.* **2005**, *350*, 476.
(165) Gavutis, M.; Jaks, E.; Lamken, P.; Piehler, J. *Biophys. J.* **2006**, *90*, 3345.
(166) Sako, Y.; Minoghchi, S.; Yanagida, T. *Nat. Cell Biol.* **2000**, *2*, 168.
(167) Zhou, H.-X.; Rivas, G.; Minton, A. P. *Annu. Rev. Biophys.* **2008**, *37*, 375.
(168) Hall, D.; Minton, A. P. *Biochim. Biophys. Acta* **2003**, *1649*, 127.
(169) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739.
(170) Swaminathan, R.; Hoang, C. P.; Verkman, A. S. *Biophys. J.* **1997**, *72*, 1900.
(171) Lavalette, D.; Hink, M. A.; Tourbez, M.; Tetreau, C.; Visser, A. J. *Eur. Biophys. J.* **2006**, *35*, 517.
(172) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Clarendon Press: 1988.
(173) Pullarkat, P. A.; Fernandez, P. A.; Ott, A. *Phys. Rep.* **2007**, *449*, 29.
(174) Dix, J. A.; Verkman, A. S. *Annu. Rev. Biophys.* **2008**, *37*, 247.
(175) Haggie, P. M.; Verkman, A. S. *J. Biol. Chem.* **2002**, *277*, 40782.
(176) Golding, I.; Cox, E. C. *Phys. Rev. Lett.* **2006**, *96*, 098102.
(177) Weiss, M.; Elsner, M.; Kartberg, F.; Nilsson, T. *Biophys. J.* **2004**, *87*, 3518.
(178) Odijk, T. *Biophys. J.* **2000**, *79*, 2314.
(179) Banks, D. S.; Fradin, C. *Biophys. J.* **2005**, *89*, 2960.
(180) Kuttner, Y. Y.; Kozer, N.; Segal, E.; Schreiber, G.; Haran, G. *J. Am. Chem. Soc.* **2005**, *127*, 15138.
(181) Lavalette, D.; Tetreau, C.; Tourbez, M.; Blouquit, Y. *Biophys. J.* **1999**, *76*, 2744.
(182) McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, *13*, 276.
(183) Minton, A. P. *Methods Enzymol.* **1998**, *295*, 127.
(184) Asakura, S.; Oosawa, F. *J. Chem. Phys.* **1954**, *22*, 1255.
(185) Vrij, A. *Pure Appl. Chem.* **1976**, *48*, 471.
(186) Verma, R.; Crocker, J. C.; Lubensky, T. C.; Yodh, A. G. *Phys. Rev. Lett.* **1998**, *81*, 4004.
(187) Chatterjee, A. P.; Schweizer, K. S. *J. Chem. Phys.* **1998**, *109*, 10464.
(188) Fuchs, M.; Schweizer, K. S. *J. Phys.: Condens. Matter* **2002**, *14*, R239.
(189) Atha, D. H.; Ingham, K. C. *J. Biol. Chem.* **1981**, *256*, 12108.
(190) Rivas, G.; Minton, A. P. *J. Mol. Recognit.* **2004**, *17*, 362.
(191) Rivas, G.; Fernandez, J. A.; Minton, A. P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3150.
(192) Farnum, M.; Zukoski, C. *Biophys. J.* **1999**, *76*, 2716.
(193) Kulkarni, A. M.; Chatterjee, A. P.; Schweizer, K. S.; Zukoski, C. F. *J. Chem. Phys.* **2000**, *113*, 9863.
(194) Liu, W.; Bratko, D.; Prausnitz, J. M.; Blanch, H. W. *Biophys. Chem.* **2004**, *107*, 289.
(195) Vivares, D.; Belloni, L.; Tardieu, A.; Bonnete, F. *Eur. Phys. J. E* **2002**, *9*, 15.
(196) Zhou, H.-X. *J. Mol. Recognit.* **2004**, *17*, 368.
(197) Ladurner, A. G.; Fersht, A. R. *Nat. Struct. Biol.* **1999**, *6*, 28.
(198) Kozer, N.; Schreiber, G. *J. Mol. Biol.* **2004**, *336*, 763.
(199) Schlarb-Ridley, B. G.; Mi, H.; Teale, W. D.; Meyer, V. S.; Howe, C. J.; Bendall, D. S. *Biochemistry* **2005**, *44*, 6232.
(200) de Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, 1979.
(201) Kozer, N.; Kuttner, Y. Y.; Haran, G.; Schreiber, G. *Biophys. J.* **2007**, *92*, 2139.

CR800373W