Simulation of the Diffusional Association of Barnase and Barstar

R. R. Gabdoulline and R. C. Wade
European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

ABSTRACT The rate of protein association places an upper limit on the response time due to protein interactions, which, under certain circumstances, can be diffusion-controlled. Simulations of model proteins show that diffusion-limited association rates are $\sim 10^6$–10^7 M^{-1} s^{-1} in the absence of long-range forces (Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 89:3338–3342). The measured association rates of barnase and barstar are 10^8–10^9 M^{-1} s^{-1} at 50 mM ionic strength, and depend on ionic strength (Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3:427–431), implying that their association is electrostatically facilitated. We report Brownian dynamics simulations of the diffusional association of barnase and barstar to compute association rates and their dependence on ionic strength and protein mutation. Crucial to the ability to reproduce experimental rates is the definition of encounter complex formation at the endpoint of diffusional motion. Simple definitions, such as a required root mean square (RMS) distance to the fully bound position, fail to explain the large influence of some mutations on association rates. Good agreement with experiments could be obtained if satisfaction of two intermolecular residue contacts was required for encounter complex formation. In the encounter complexes, barstar tends to be shifted from its position in the bound complex toward the guanine-binding loop on barnase.

INTRODUCTION

The association of two proteins is a ubiquitous event in biological systems that is fundamental to processes such as signal transduction, transcription, cell cycle regulation, and immune response. The speed of the association phase places an upper limit on the speed of the response resulting from protein interactions in vivo. Often, at least one of the interacting proteins is free to move in the intra or extracellular environment and must find its partner by diffusion. The association rate is limited by the time it takes to bring the "reactive patches" of the proteins together by diffusion (Calef, 1983; Berg and von Hippel, 1985). When the post-diffusional step of the association process is much faster than diffusional dissociation, the "reaction" is diffusion-controlled, i.e., the association rate of two molecules is defined by their diffusional encounter (DeLisi, 1980). Clearly, determination of whether a reaction or response is diffusion-controlled is crucial to protein design studies, since the means by which molecular properties are altered will depend on whether the rate-limiting step is a diffusion or not.

The kinetics of protein-protein interactions have recently become the target of numerous experimental measurements (Szabo et al., 1995). To some extent this is due to the commercial availability (since 1990) of instruments based on surface plasmon resonance (Malmqvist, 1993; Szabo et al., 1995). These, in principle, provide a generally applicable noninvasive method to measure the kinetics of protein-protein interactions (without the necessity for labeling). Reported experimental values of association rates cover a wide range, from 10^3 to 10^9 M^{-1} s^{-1}. The lower limit is largely dictated by experimental limits on the measurable rates (Szabo et al., 1995), while the upper limit is based on measured association rates in solution for thrombin-hirudin (Stone et al., 1986) and barnase-barstar (Schreiber and Fersht, 1996) association. Diffusion control of protein association is indicated by the following five properties of association rates: 1) fast ($\geq 10^6$ M^{-1} s^{-1}) under typical conditions; 2) inverse dependence on the viscosity of the solvent (Calef, 1983) implying linear dependence on the diffusion constant; 3) dependence on the ionic strength of the solution indicating the importance of the long-range electrostatic forces; 4) their temperature dependence is governed by that of the solvent viscosity. Diffusion constants of proteins in water increase by a factor of ~ 2 over the temperature range 288–313K (Gosting, 1956), and this causes diffusion-controlled rates to increase by the same factor, as has been observed for some antibody-antigen complexes (Johnstone et al., 1990; Ito et al., 1995). 5) sensitivity to the diffusional environment, e.g., whether association takes place with both proteins free to diffuse in solution or with one of the proteins immobilized on a surface.

None of these properties alone or in combination is a proof of diffusional control. Therefore, detailed simulation of diffusional association is a necessary step in defining the mechanism of interaction.

The interaction of barnase, a ribonuclease that acts extracellularly, with its intracellular inhibitor, barstar (produced by *Bacillus amyloliquefaciens*), is among the stron-

Received for publication 30 December 1996 and in final form 4 February 1997.

Address reprint requests to Dr. Rebecca C. Wade, European Molecular Biology Laboratory, Meyerhofstraße 1, D-69117 Heidelberg, Germany. Tel.: 49-6221-387-553; Fax: 49-6221-387-517; E-mail: wade@embl-heidelberg.de.

0006-3495/97/05/1917/13 $2.00

gest known interactions between proteins (Hartley, 1993; Schreiber and Fersht, 1993), and it has been the target of a large number of experimental studies. The interactions of the proteins have been extensively studied by single point mutations and double mutant cycles (Schreiber and Fersht, 1995) and a system has been constructed to perform in vivo selection to find compensating mutations (Jucovic and Hartley, 1996). The kinetic parameters of barnase-barstar interactions have been measured by Schreiber and Fersht (1995) using the stopped-flow fluorescence technique. The rates for association of the wild-type proteins are fast (Schreiber and Fersht, 1995), and viscosity-dependent (Schreiber and Fersht, 1996), suggesting that their association is diffusion-controlled. The importance of electrostatic interactions in determining the association rates is indicated by their ionic strength dependence and sensitivity to protein mutation (Schreiber and Fersht, 1996). Double mutant cycle analysis points to the rate-limiting formation of an early, weakly specific, encounter complex dominated by electrostatic interactions, which then goes on to dock and form a high-affinity complex. The measured association rates for 10 different single-residue mutants and one double-residue mutant analyzed in the present study (see Table 1), range from $3 \cdot 10^7$ to $2 \cdot 10^9$ M^{-1} s^{-1} in the presence of 50 mM NaCl, pH 8.0, and temperature 25°C (Schreiber and Fersht, 1996).

It was recently recognized (Northrup and Erickson, 1992) that rates of interprotein association governed purely by diffusion may be on the order of $2 \cdot 10^6$ M^{-1} s^{-1}, which is close to the experimentally measured rates for antigen-antibody association (see, for example, Ward et al., 1989; Ito et al., 1995). This theoretically derived rate is half the diffusional rate for aligning two of four contact pairs on model spherical proteins within 2 Å. During Brownian dynamics simulations of the association of these model proteins, 50% of the encounter complexes having two aligned contacts evolve to complexes with three aligned contacts if reasonable short-range attractive forces are taken into account (Northrup and Erickson, 1992). Brownian dynamics simulations with detailed protein models have previously been successfully applied to study diffusion-controlled enzymes (see Madura et al., 1995; Wade, 1996, and references therein) to explore the dependence of enzyme-substrate encounter rates on the enzyme's electrostatic potential and dynamics, e.g., of active site loops (Wade et al., 1994). Such Brownian dynamics simulations have also been used to study the association of proteins, viz: antibody-antigen association (Kozack and Subramaniam, 1993; Kozack et al., 1995), cytochrome interactions to model bimolecular electron transfer (Northrup et al., 1988; Northrup et al., 1993), and bovine pancreatic trypsin inhibitor interactions to examine disproportionation rates for surface disulfide groups (Nambi et al., 1991).

The goal of the present study is to reproduce the experimental data for barnase-barstar association using Brownian dynamics simulations, and thereby gain insight into diffusional pathways and the structural and kinetic properties of diffusional encounter complexes formed by barnase and barstar and their mutants. The extensive experimental characterization of this system also enables us to use it as a test system for calibrating the simulation methodology for making predictions of association rates for new mutants and other proteins. In addition, the Brownian dynamics simulations provide a test of whether three-dimensional models of mutant proteins, constructed when their structure has not been determined experimentally, are of sufficient quality to compute accurate kinetic data.

The main assumptions invoked for simulating barnase and barstar (wild-type and mutants) are that 1) association is diffusion-controlled, and 2) the criteria defining diffusional encounters are the same for all mutants. Thus, we evaluate whether differences in association rates of almost two orders of magnitude on making one to two point mutations can result solely from differences in the diffusional encounter rates of the two proteins. In the simulations, we neglect protein flexibility, describe long-range interactions by a continuum electrostatics model, model van der Waals repulsion by exclusion forces, and neglect hydrodynamic and short-range forces. The simulation results show that, even within the limits of the model employed, the experimental rates can essentially be recovered by computer simulations of diffusional encounter, confirming the validity of assumptions 1) and 2).

In the Methods section, we describe the Brownian dynamics algorithm. It is similar to that used by Northrup et al. (1987; 1988) in simulations of the cytochrome *c*-cytochrome *c* peroxidase electron transfer system, except for differences in the method of computing interaction energies between macromolecules and defining encounter complex formation. We consider three different definitions of formation of the interprotein encounter complex (the final point in the diffusional association of the two proteins) and show that one gives results in good agreement with experimental results. The computed rates are presented in the Results section. In the Discussion section, we compare experimental and simulation data, and describe an association pathway for the two proteins consistent with these data.

METHODS

Computation of association rates by Brownian dynamics simulation

To compute association rates, many trajectories are simulated that start with the proteins at uniformly random orientations at a center-to-center separation distance b, and finish when the protein centers reach a separation distance $c > b$. If a fraction β of the total number of trajectories satisfies a defined reaction criterion, then the reaction rate constant k can be derived (Northrup et al., 1983) (see also Allison et al., 1986; Zhou, 1990; Northrup, 1994) by applying the formula

$$k = \frac{k(b) \cdot \beta}{1 - (1 - \beta) \cdot k(b)/k(c)} \quad (1)$$

where $k(x)$ is the rate constant for diffusion to a relative separation x. If, as is assumed here, the proteins do not experience interaction forces at separations $\geq b$, then $k(x)$ is given by the well-known Smoluchowski

expression for a two-sphere diffusional encounter, where D is the translational diffusion coefficient for relative motion (Smoluchowski, 1917).

The diffusion equation is solved with the Ermak-McCammon algorithm (Ermak and McCammon, 1978). The translational Brownian motion of two interacting proteins is simulated as the displacement $\Delta\mathbf{r}$ of the relative separation vector $\mathbf{r}$ in a time step Δt according to the relation

$$\Delta\mathbf{r} = \frac{D \cdot \Delta t}{k_B T} \cdot \mathbf{F} + \mathbf{S} \qquad (2)$$

where $\mathbf{F}$ is the systematic interparticle force, $k_B T$ is the Boltzmann constant times temperature, and $\mathbf{S}$ is the stochastic component of displacement caused by collisions of the proteins with solvent molecules. Each component of $\mathbf{S}$ is generated as normally distributed random numbers with zero mean, and the variance $\langle S^2 \rangle = 6D\Delta t$. Two analogous formulas are used to generate the rotational motions of the proteins in terms of rotation angle $\mathbf{w}_i = (w_{1i}, w_{2i}, w_{3i})$, torque $\mathbf{T}_{ij}$ acting on protein i due to protein j, and rotational diffusion constant D_{iR} of each protein i ($i, j = 1, 2, i \neq j$):

$$\Delta\mathbf{w}_i = \frac{D_{iR} \cdot \Delta t}{k_B T} \cdot \mathbf{T}_{ij} + \mathbf{W}_i, \qquad (3)$$

where $\mathbf{W}_i$ is a stochastic term with $\langle \mathbf{W}_i \rangle = 0$ and $\langle \mathbf{W}_i^2 \rangle = 2D_{iR}\Delta t$. The diffusional properties of the molecules are thus assumed to be isotropic. Neither hydrodynamic steering effects (Brune and Kim, 1994) nor their coupling to electrostatic steering (Antosiewicz and McCammon, 1995; Antosiewicz et al., 1996) are considered. These may contribute to the absolute values of the association rates in our system. However, it can be expected that their contribution to the differences in the association rates between protein mutants and the dependence of association rates on the solvent's ionic strength is small.

Checks for multiple reaction conditions are performed during each simulation. As described in McCammon et al. (1986), the reaction events are counted at their first occurrence and do not influence the motions of the proteins.

Computation of forces

The computation of forces and torques follows the algorithm described in Northrup et al. (1988), with the main difference being that we use potential derived charges (Gabdoulline and Wade, 1996) for the proteins instead of test charges. Namely, a full partial atomic charge model of each protein is used to compute each protein's electrostatic potential separately by numerical solution of the finite difference Poisson-Boltzmann equation (Davis and McCammon, 1989), taking into account the inhomogeneous dielectric medium and the surrounding ionic solvent. "Effective charges" are then computed for each protein by fitting to its potential. The array of effective charges for protein 2 (1) is placed on the electrostatic potential grid of protein 1 (2) to compute the forces and torques acting on protein 2 (1). According to our previous study (Gabdoulline and Wade, 1996), the above algorithm for computing forces gives a good approximation to the forces derived by solving the finite difference equation in the presence of both proteins, unless the separation of the protein surfaces is less than twice the solvent probe radius and the desolvation energies of the charges in one protein due to the low dielectric cavity of the other protein become significant.

For comparison, we also examined the use of test charges for the computation of forces. The forces were computed treating both proteins symmetrically, and therefore the interactions for this test charge model are different from those in Northrup et al. (1988), where the forces acting on one of the proteins were computed using only two charges (dipolar treatment).

Short-range repulsive forces are treated by an exclusion volume prohibiting van der Waals overlap of the proteins. If a move during a time step would result in van der Waals overlap, the Brownian dynamics step is repeated with different random numbers until it does not cause an overlap.

Materials and parameters

The experimental association rates are listed in Table 1. The modeled ionic strength values correspond to the experimental ionic strength of the solvent due to addition of NaCl, at pH 8 and 25°C (Schreiber and Fersht, 1996). Rates were computed for the wild-type proteins and 11 of the 19 mutants studied by Schreiber and Fersht (1996) chosen to cover the range of rates observed, and omitting those mutants with rates very close to the wild-type proteins.

The 2-Å resolution x-ray coordinates of the barnase-barstar complex (Buckle et al., 1994) were extracted from the Brookhaven protein data bank (Bernstein et al., 1977) (file identifier: 1brs). The A (for barnase) and D (for barstar) chains were used. Two other pairs of chains in the crystal structure were used to model the (few) missing side chain atoms in the A and D chains. Crystallographic water molecules were removed. Polar hydrogen atoms were added and their positions optimized by energy minimization with the CHARMM program (Brooks et al., 1983) using the QUANTA molecular graphics package (QUANTA, 1992). Side-chain conformations were kept the same as in the barnase-barstar complex. Mutations were modeled, either by merely deleting the side chain for the alanine mutations or by modeling in the new side chain for the *D22M*, *E73W*, and *R83Q* mutants.

The UHBD program (Madura et al., 1994; 1995) was used to calculate the electrostatic potentials of the proteins given by the linearized Poisson-Boltzmann equation. Partial atomic charges and atomic radii were assigned from the OPLS parameter set (Jorgensen and Tirado-Rives, 1988). The protonation states of titratable residues were assigned according to their standard protonation states at the experimental pH of 8.0 (Schreiber and Fersht, 1995). Dielectric constants were assigned 2 in the protein interior, defined as the region within the van der Waals surface, and 78 in the exterior. Dielectric boundary smoothing (Davis and McCammon, 1990) was used in all finite difference calculations. Solution ionic strengths of 50, 150, 300, and 500 mM were modeled using the Debye-Hückel description of the solvent outside a 2-Å probe-accessible surface of the proteins. Grids with dimensions of 149 × 149 × 149 points and a 1.0-Å spacing centered on each of the proteins were used to represent their electrostatic potentials. Consequently, the interaction forces between one of the proteins and the charges on the other were truncated at 74 Å from the center of the first protein. The extents of the two proteins are such that none of the atoms of barnase is further than 23 Å from the center of the protein, the corresponding distance for barstar is 22 Å; i.e., the interactions are truncated >49 Å from the protein surfaces. The influence of such a truncation of potentials on the computed association rates was estimated by control calculations

TABLE 1 Experimental association rates compared with computed rates

Mutations	Exp. rate	Computed d_C = 5.5 Å	3 × Exp. rate	Computed d_C = 6.25 Å
None (WT)	2.86	4.6 (0.3)	8.58	10 (0.6)
D22M	4.65	6.1 (0.6)	14.0	14 (0.3)
K27A	0.49	0.67 (0.1)	1.49	1.4 (0.2)
D54A	6.87	22 (3.0)	20.6	33 (2.0)
R59A	0.31	0.24 (0.07)	0.92	0.81 (0.1)
E60A	13.9	29 (1.6)	41.7	36 (2.0)
E73W	6.82	30 (1.5)	20.5	35 (1.3)
R83Q	1.01	0.85 (0.2)	3.03	2.8 (0.5)
R87A	1.32	1.3 (0.26)	3.96	3.6 (0.5)
*E76A**	1.30	0.85 (0.3)	3.90	2.8 (0.6)
*E80A**	1.50	1.7 (0.3)	4.50	3.5 (0.8)
D54A + E60A	18.7	48 (2.5)	56.1	56 (2.5)

50 mM NaCl, 5 mM Tris-HCl buffer, pH 8.0, 25°C (Schreiber and Fersht, 1996) rates are given in units of $10^8\ M^{-1} \cdot s^{-1}$. The standard error of the mean for experimental association rate constants is 12% (Schreiber and Fersht, 1995). Standard deviations in computed rates are given in brackets. *Mutations in barstar; otherwise, the mutated residue is in barnase.

using a smaller 109 × 109 × 109 point grid with a 1.0 Å spacing. The relative rates at 50 mM ionic strength for different mutants compared to wild-type proteins changed negligibly when this smaller grid was used. Worth noting, though, is that the absolute values at 50 mM were 25% smaller, indicating small but non-negligible electrostatic steering due to forces between atoms further than 30 Å apart.

The effective charges (Gabdoulline and Wade, 1996) for each protein were fit to reproduce the electrostatic potential in a layer 3 Å thick extending outward from the protein's accessible surface computed as defined by a 4-Å radius probe. The charges were placed on the carboxylate oxygens of Asp, Glu, and the C-terminus, and the amine nitrogens of Lys, Arg, and the N-terminus. Forty-seven effective charges were assigned for wild-type barnase and 45 for wild-type barstar.

To model exclusion forces, either 1.0- or 0.5-Å spacing grids attached to one of the proteins were used. A grid point was assigned a value of 1 if it was within the probe-accessible surface of the protein, and a value of 0 otherwise. Overlap was defined to happen when any solvent-accessible atom of the second protein was projected onto a grid point having a value of 1. The procedure is the same as the one used by Northrup et al. (1988). It implies that all the atoms of the second protein have a radius equal to the probe radius used to define the probe-accessible surface of the first protein. We used a probe radius of 2.0 Å for consistency with the ion exclusion radius value of 2.0 Å in the Debye-Hückel treatment of ionic strength. The spacing of the exclusion grid did not detectably influence the computed rates, although a systematic decrease in encounter times (not shown) was noticeable upon decreasing the grid spacing to 0.5 Å.

A relative translational diffusion constant of 0.030 Å^2/ps was used, based on the individual diffusion constants of 0.015 Å^2/ps assigned to both barnase and barstar. Rotational diffusional constants of $4.0 \cdot 10^{-5}$ and $4.5 \cdot 10^{-5}$ radian^2/ps were assigned to barnase and barstar, respectively, in aqueous solution at 25°C. The Einstein-Stokes relationship gives these constants for spheres having radii of ~16 Å. This radius corresponds to the effective radii of the proteins computed as the distance from the center to where the atomic density is ⅔ the density at the center of the protein. The translational diffusion constants assigned compare favorably with the diffusion constants of similarly sized proteins (see Creighton, 1993).

Brownian dynamics trajectories were started at an intermolecular separation b = 100 Å. At this distance, there is no interaction between the proteins because of the electrostatic potential cutoff applied. The trajectory truncation distance, c, was $5b$ = 500 Å. It was chosen so that doubling it did not change the computed rates within statistical errors. A variable time step was used. It was increased linearly with intermolecular separation, being 2.0 ps when two molecules were within 5 Å from overlap. These time steps correspond to a mean random displacement of 0.6 Å when the proteins are close and 4.5 Å at the greatest protein separations. We considered smaller (by a factor of 4) time steps to ensure that the results were not influenced by this computational parameter.

The number of trajectories, N, run to estimate an association rate varied from 10,000 for wild-type proteins at 50 mM ionic strength up to 400,000 for cases for which rates were small, and hence more simulations were required to achieve reliable rate estimates. Statistical errors were estimated by two different methods: 1) assuming a Poisson distribution of the encounter event probabilities (Pollard, 1976), and using a simple estimate of statistical error as $K^{-1/2} \cdot 100\%$, where K is the number of encounter events; multiplication of this error by 1.65 gives a 90% confidence level (Madura et al., 1995); or 2) by performing four different Brownian dynamics runs and deriving the error as the RMSD of the four rates from their average. The errors are similar for the two methods, and those derived by the latter method are shown in the error bars in the figures and Table 1. In our simulations, rates of $3 \cdot 10^8$ M^{-1} s^{-1} correspond to encounter probabilities, β, of $2 \cdot 10^{-3}$. Thus, the number of encounter events in 10,000 trajectories is 200, giving an error estimate of 7%. We ran more trajectories for the cases with smaller rates to ensure comparable error estimates. It was not possible to perform simulations with no electrostatic forces at this error level or even with errors of <100%, and we used a maximum of 400,000 trajectories to estimate the trend in the dependence of the rate on encounter distances. The simulation of 10,000 trajectories took 2–12 h of CPU time on an SGI Power Challenge computer (single R8000 processor), depending on the simulation conditions (ionic strength value, encounter criteria, protein identities).

Encounter definitions

Three different criteria for defining formation of a diffusional encounter complex were considered. In all cases, we define one parameter measuring the closeness of the two proteins and require this to reach a certain boundary value for encounter complex formation. Then we compute the rates of diffusion to a set of boundaries defined by different values of this parameter. By equating experimentally measured rates to the computed rates, we derive the value of the parameter. The derived value shows whether the criterion leads to realistic distances. A further test is to compare the thus-derived boundary parameters under different conditions, i.e., different interaction forces due to residue mutations and changing the salt concentration. We assume that the encounter distances remain the same in all cases, which implies that the encounter parameters are not modulated by point mutations or salt conditions.

The first criterion is geometric. It depends on the root mean square (RMS) distance, d, between the coordinates of two atoms in barstar ($\mathbf{x}_i$) during a simulation and the coordinates of these atoms in the crystallographically observed complex of the two proteins ($\mathbf{x}_{i0}$):

$$d = \sqrt{\tfrac{1}{2} \cdot (|\mathbf{x}_1 - \mathbf{x}_{10}|^2 + |\mathbf{x}_2 - \mathbf{x}_{20}|^2)} \qquad (4)$$

The $C\gamma$ atoms of Asp-35 (atom i = 1) and Asp-39 (i = 2) of barstar were chosen for this definition, as these residues are involved in many interactions with barnase residues in the barnase-barstar complex. The distance d is a measure of the degree of closeness to the target position. Consequently, it will depend on the choice of atoms chosen to monitor the distance. The same RMS distance value as derived for the barnase-barstar case can only be used for other proteins if they have analogous geometries.

The second criterion uses electrostatic interaction energy, E_{el}. E_{el} is computed using the same potentials and charges as used for computing the steering forces for Brownian motion. Note that because effective charges are used, interactions at contact distances are not described well. We find that the use of this criterion leads to overestimates of steering, as would be expected unless rate enhancement is not of electrostatic origin.

The third criterion is the number of correct residue-residue contacts between the two proteins that are shorter than a defined length (d_C). A list of contacts to be monitored is derived from analyzing H-bonds observed in the crystal structure of the complex (Buckle et al., 1994). It is important to eliminate mutually dependent H-bonds. For example, in the case when two atoms of one residue make H-bonds with two atoms of another, we retain only the contact between the most accessible atoms. The reason for this elimination is to construct mutually independent contacts that make approximately additive contributions to the short-range interaction energy. The contact pairs we defined are listed in Table 2 and shown in Fig. 1, together with residues mutated in barnase and barstar. The contact distance d_C and the number of contacts (1, 2, or 3) required to be met were chosen

TABLE 2 Atom pairs (contacts) used to monitor encounter complex formation by the third criterion described in Methods

Barstar	Barnase	Contact distance in crystal structure of complex (Å)
OE1 Glu-76	NH2 Arg-59	3.0
ND2 Asn-33	O His102	3.2
N Leu-34	OE2 Glu-60	2.8
OD1 Asp-35	N Arg-59	2.9
OG1 Thr-42	NZ Lys-27	3.0
OH Tyr-29	O Arg-83	2.7
OD1 Asp-39	NH2 Arg-83	2.5
O Gly-43	NH1 Arg-83	3.1

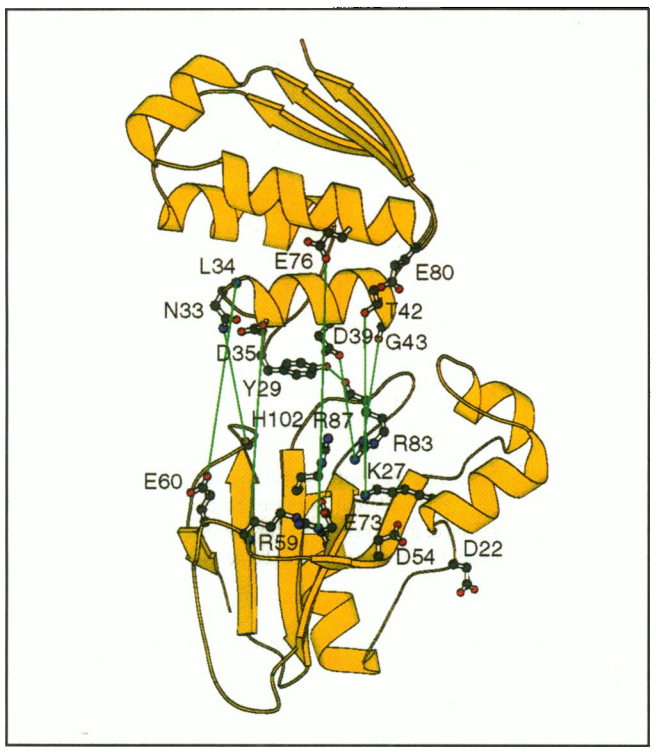

FIGURE 1 Location of the mutated residues listed in Table 1 and the interacting (contact) atom pairs listed in Table 2 of barnase and barstar. The atom pairs are connected by green lines. Barnase (*bottom*) and barstar (*top*) are shown in ribbon representation with the structures observed in the crystal structure of their complex. To assist visualization, the proteins were pulled 12.5 Å apart and rotated around their centers by −60° (barstar) and +15° (barnase) from their positions in the crystal structure complex. This figure was generated using the MolScript program (Kraulis, 1991).

so that rates computed from the simulations agree with experimentally measured values.

The number of possible contacts can vary with protein mutation. For example, if a residue with an atom that is part of a listed contact pair is mutated to alanine, the contact is eliminated. On the other hand, when modeling an Arg to Gln mutation for an Arg involved in a contact pair, the identity of the atom defining the contact pair is modified.

A more detailed contact criterion was also investigated. All possible donor-acceptor atom pairs with one atom in barnase and the other <6 Å away in barstar in the crystallographically observed complex were tabulated. These contacts were then clustered into eight groups. During a simulation the number of contacts in different groups was counted.

RESULTS

Barnase-barstar interaction energies

The long-range electrostatic interactions between barnase and barstar favor orientations close to those in their crystallographically observed complex. This is apparent from computation of the interaction energies for many different relative orientations of the two proteins. This was done by fixing one of the proteins and placing the other at many random orientations around the first, with its center scanning the surrounding 3D space and, at each position, computing the interaction energy. From this energy, a distribution function for the second protein around the fixed protein was computed using the expression:

$$\rho(\mathbf{x}) = \frac{1}{N} \cdot \sum_{i=1}^{N} e^{-E(\mathbf{x},i)/k_B T} \tag{5}$$

Here, $\mathbf{x}$ denotes the Cartesian coordinates of the center of the second protein and the summation is over the N random angular orientations sampled with the center of the second protein at $\mathbf{x}$. k_B is the Boltzmann constant and T is the temperature. $E(\mathbf{x}, i)$ is the interaction energy at orientation i and is given by the electrostatic interaction free energy if there is no overlap, and set at infinity if there is. This distribution function is a projection of the distribution functions of two proteins in six dimensions onto three dimensions, and is therefore a Boltzmann distribution of the center of the second protein around the first, when only exclusion and electrostatic forces are present. Fig. 2 shows contours of $\rho(\mathbf{x})$ when the first (fixed) protein is barnase and the second is barstar (Fig. 2 *A*) and vice versa (Fig. 2 *B*). These show that the interactions tend to be most favorable when both proteins are near their orientation in the crystallographic complex, although there is an additional favorable position for barstar around barnase, which shifts barstar's average electrostatically favored position toward the third loop (residues 56–69) of barnase.

Brownian dynamics simulations

Results are presented for simulations with each of the three criteria for formation of a diffusional encounter complex between barnase and barstar.

RMS distance criterion

Fig. 3 shows how the ratio of the computed to the experimental rates (at 50 mM ionic strength) varies with RMS distance, d. For the wild-type proteins, the computed rate is equal to the experimental rate when d = 6.5 Å. In the crystallographic complex d = 0.0 Å, so this distance is too large to be a realistic measure of the distance at which the short-range interactions between the proteins (omitted from our model) become strong enough to ensure that complexation occurs subsequent to formation of the diffusional encounter complex. At d = 6.5 Å, it is likely that other residues in barstar, in addition to the two used in the criterion definition, interact significantly with barnase. In the crystal structure, the criterion-defining residues in barstar are positioned near the middle of the concave binding face on barnase. Thus, even at large values of d, the criterion-defining residues in barstar should be able to interact with residues toward the edge of barnase's concave surface.

While the computed rates for some mutants (e.g., *E60A*, *K27A*) approximately reproduce the experimental rates using the same value of d as for the wild-type proteins, this is

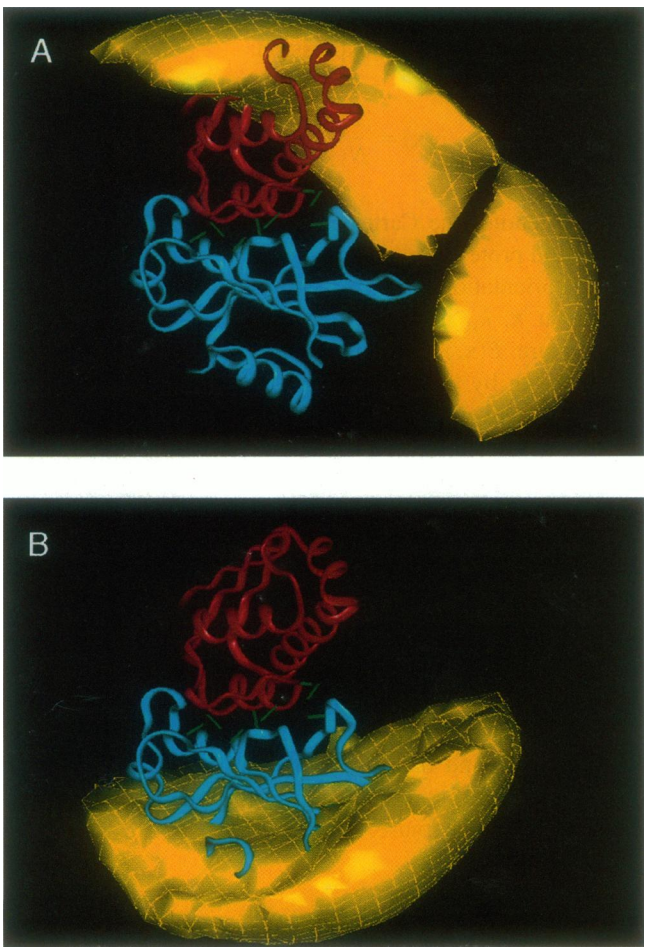

FIGURE 2 Distribution functions, $\rho(\mathbf{x})$, of (*A*) the center of barstar around barnase and (*B*) the center of barnase around barstar. $\rho(\mathbf{x})$ is given by Eq. (5) and computed considering electrostatic interaction and exclusion energies only. The contours (*yellow*) are at $\rho(\mathbf{x}) = 2$ (which corresponds to an affinity of ½ M). Barnase is shown in blue and barstar in red. The interfacial contact pairs (listed in Table 2) are shown in green. Note that in this figure, barnase is rotated by ~135° around the *z*-axis from its position in Fig. 1, so that loop 3 (residues *56–69*) is moved from the front to the right-hand side.

not the case for all mutants. Exceptions are the *D54A* and *R59A* mutants whose rates are overestimated by a factor of 3 at d = 6.5 Å. Agreement with the experimentally measured rates is obtained for these mutants at d = 5.0–5.5 Å (at which distance, agreement with experiment is also good for the *E60A* mutant).

Electrostatic energy criterion

The value of E_{el} at which computed and experimental rates for the wild-type proteins match is $-12\ k_BT$ or -7.3 kcal/mol. The values of E_{el} at which computed and experimental rates match for the three mutants examined vary between -15 and $-8.5\ k_BT$ (-9 and -5 kcal/mol) (Fig. 4). Thus, it is impossible to compute the rates with any accuracy using a single electrostatic energy value as the criterion for encounter complex formation. This can be understood by

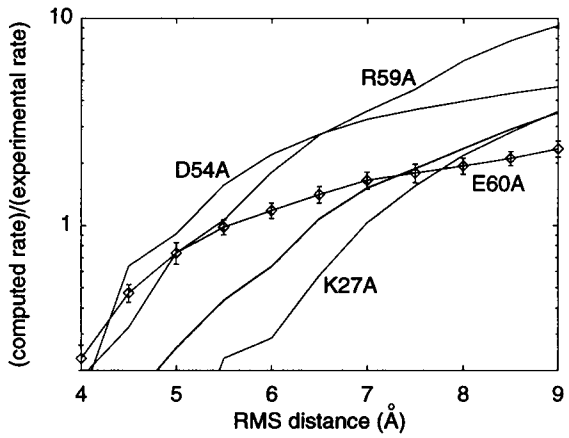

FIGURE 3 The ratio of computed to experimental rates for wild-type (in bold) barnase and barstar and four barnase mutants (labeled) is shown versus the RMS distance, d, (given in Å) for the two-atom RMS distance criterion. (Experimental rates are given in Table 1).

considering the *K27A* mutant. The decrease in rate of this mutant compared to wild-type is caused by the decrease in attractive electrostatic forces. With this criterion, the decrease in rate derives not only from the decrease in the electrostatic steering force but also from the decrease in the size of the reaction surface for encounter complex formation, because the interaction energies are lower due to the mutation. Consequently, the electrostatic interaction energy does not define the encounter complex, the definition of which is also dependent on short-range interactions which are not fully modeled.

Contact pair criteria

When satisfaction of only one contact is required for diffusional encounter complex formation, the computed rates are only close to the experimental ones at very short contact distances of ~3.5 Å (Fig. 5), which is the minimum distance

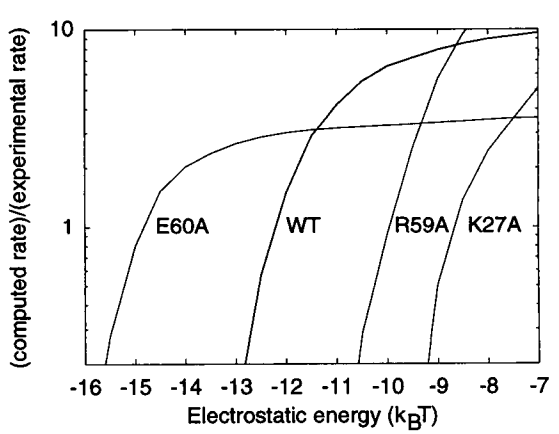

FIGURE 4 The ratio of computed to experimental rates for wild-type (in bold) and three barnase mutants is shown versus electrostatic interaction energy, E_{el}, (given in units of k_BT) when this is used as the criterion for diffusional encounter complex formation.

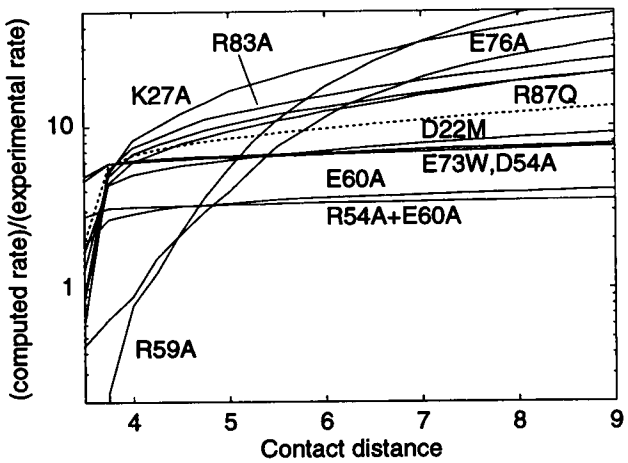

FIGURE 5 The ratio of computed to experimental rates for wild-type (*dotted line*) and 11 mutant proteins (*continuous lines*), calculated assuming that only one contact is required for encounter complex formation, is shown versus the contact distance in Å that must be satisfied.

allowed between the two atoms in our model. Although the ionic strength dependence satisfactorily fits the experimental dependence (not shown), the rates computed for the different mutants with this contact distance criterion clearly do not correlate with experimental data. The removal of the most exposed and the most frequently appearing contact *R59-E76* by making an *R59A* mutation results in a computed association rate about a factor of 10 lower than the experimental value for this mutant when the distance criterion is 3.5 Å.

When three contacts are required for encounter complex formation, the majority of the computed rates are close to the experimentally measured rates when the contacts are defined by a maximum distance of 7–8 Å (Fig. 6). For 6 of the 10 single-mutant proteins considered, there is a high degree of correlation between the ratio of computed to experimental rates and the contact distance: at contact distances >7.5 Å, all ratios are within ±75% of the average ratio. The outliers are: *E60A*, *D54A*, *E73A*, and *R59A*. The results for *E60A* only correlate with those of the majority of the mutants at contact distances of 7.5–9.0 Å. The rates computed for *D54A* and *E73A* are mostly overestimates and correlate with those of most mutants only at distances >9.0 Å. Compared to the majority of mutants, the rates computed for *R59A* overestimate experimental ones. The double mutant *D54A+E60A* is also an outlier which displays similar behavior to *D54A* at close contact distances, and to *E60A* at larger contact distances.

When only two contacts are required for encounter complex formation, the best fit to the experimental data is obtained if the contact distance is defined at 5.5 Å (Fig. 7). The distance 5.5 Å can be interpreted as a direct, rather than solvent-separated, interaction distance between atoms. Four mutants are outliers: *E60A*, *D54A* + *E60A*, *E54A*, and *D73A*. The first two correlate with the majority of mutants at encounter distances of 5.75–6.75 Å, and the last two at distances >6.75 Å. The behavior of the curves for these four mutants shows a marked saturation tendency at distances >6 Å, indicating that association at these distances is essentially independent of the encounter distance.

When using this two-contact criterion, the computed ionic strength dependence of the rates between 50 and 500 mM ionic strength essentially reproduces the experimental dependence (Fig. 8). As has been demonstrated before for the cytochrome *c*-cytochrome *c* peroxidase system (Zhou, 1993), the proper treatment of the interaction energies is important for reproducing the ionic strength dependence of the rates. Fig. 9 shows the ionic strength dependence of the association rates computed employing test charges to model the electrostatic interactions. The use of test charges to compute interaction energies leads to an underestimation of the steering forces. Thus, while the experimentally observed 18-fold drop in association rate on changing the ionic strength from 50 to 500 mM is modeled correctly with effective charges (Fig. 8), it is modeled as only a 10-fold

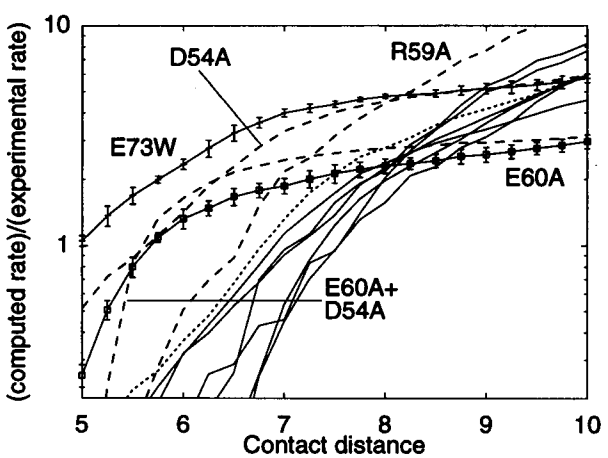

FIGURE 6 Same as Fig. 5 but for the case when three contacts are required for encounter complex formation. Mutants that are outliers to the dominant trend are shown either by dashed lines or with error bars.

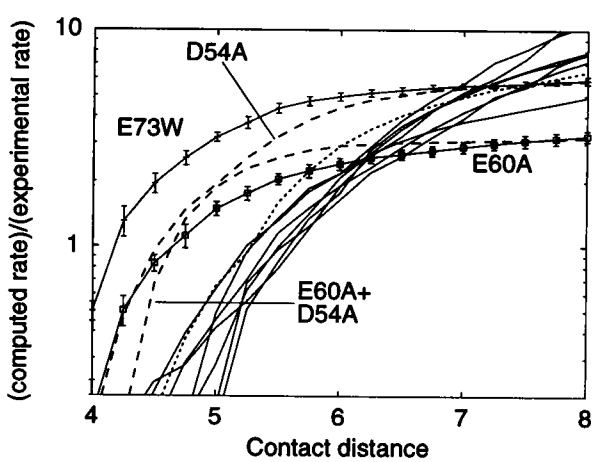

FIGURE 7 Same as Fig. 5 but for the case when two contacts are required for encounter complex formation.

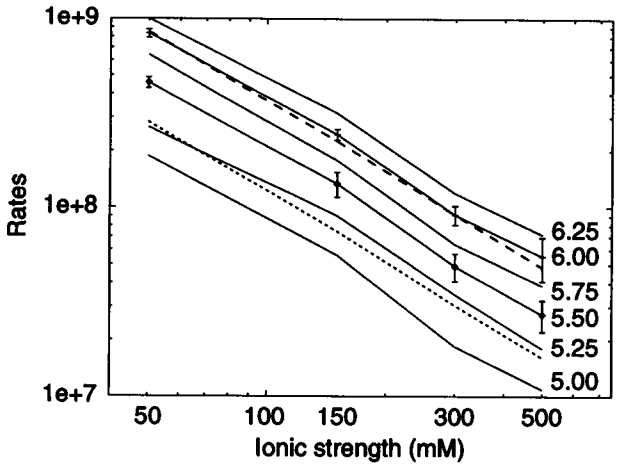

FIGURE 8 The ionic strength dependence of the rates computed using two contact atoms to determine encounter complex formation is shown for contact distances from 5 to 6.25 Å. Rates were only computed at ionic strengths of 50, 150, 300, and 500 mM. The ionic strength dependence of the experimental rate and the rate multiplied by 3 are shown by dotted and dashed lines, respectively. Note the logarithmic scales on the x- and y-axes.

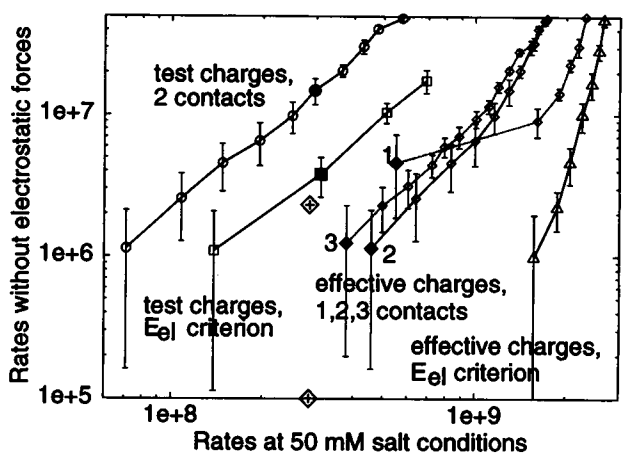

FIGURE 10 Rates computed from simulations done without electrostatic forces are shown versus the rates computed under 50-mM salt conditions. The same encounter parameter is used to compute the rates given by any point on the graph. The rates are obtained using: test charges and the two-contact criterion (*circles*); test charges and the electrostatic energy criterion (*boxes*); effective charges and 1, 2, or 3 contact criterion (*diamonds*, the curves are numbered by the number of contacts required); effective charges and the electrostatic energy criterion (*triangles*). The points with filled symbols correspond to the value of each parameter giving best agreement with experiment at 50 mM ionic strength. (No filled symbol is shown for effective charges with the electrostatic energy criterion, because with a criterion of 12 kT, the rate in the absence of forces was too slow for reactions to be sampled during simulations). The experimental data are shown by diamond+cross-signs. The smaller value is the rate derived from extrapolation (Schreiber and Fersht, 1996) to infinite ionic strength versus the rate measured at 50 mM. The larger value is the smallest rate measured (at 2000 mM) versus the rate measured at 50 mM.

drop with test charges (Fig. 9). The computed rate change on going from 300 to 500 mM is significantly smaller than that measured experimentally, implying that the test charge interactions at these (and higher) ionic strengths are very small at the encounter contact distances (6–6.5 Å) derived to fit the experimentally measured association rate at 50 mM salt. The use of effective charges to compute forces reproduces the ionic strength dependence at ionic strengths up to 500 mM, and models the significant steering forces that are present even at such high ionic strengths. Consequently, the measured ionic strength dependence of the rates may be ascribed solely to changes in the steering forces. The association mechanism is the same at salt concentrations ranging from 50 to 500 mM.

Fig. 10 shows the relationship between the rates computed with and without invoking electrostatic interactions, when using the same encounter criterion parameter. Four points should be noted: 1) The enhancement of the rate by electrostatic forces is greater when effective charges are used than when test charges are used for both types of reaction criteria shown. 2) For both test and effective charge models, the enhancement of the rate by electrostatic forces is greater when the electrostatic energy criterion is used than when the contacts criterion is used. 3) Only with effective charges and a reaction criterion requiring two or three contacts are the ratios between rates at 50 mM ionic strength and rates without electrostatic forces consistent with the ratios derived directly from experimental data. 4) When the two-contact criterion is applied with the optimum contact distance of 5.5 Å using the effective charges model, the rate at 50 mM ionic strength is enhanced over that in the absence of electrostatic forces by a factor of ~450.

Further improvements in the contacts model, when the residue-residue contacts were presented as a set of atom-atom contacts, resulted in little improvement (not shown) over the correlation with experimental data already obtained. This is probably because of limitations of the protein model at short distances: the proteins are treated as rigid, and short-range interactions are incompletely modeled. At short distances, the mobility of the hydrogen-bond network

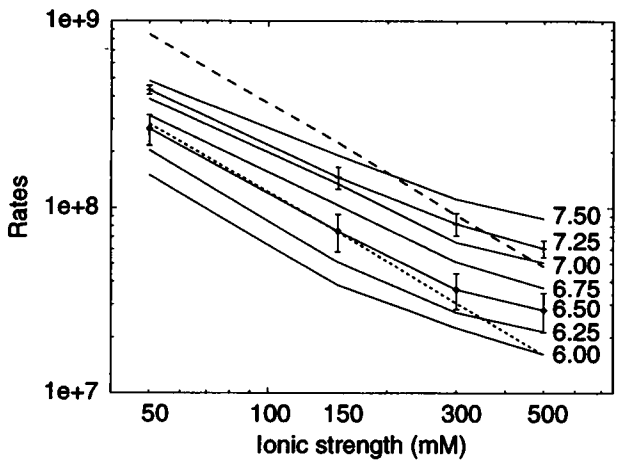

FIGURE 9 Same as Fig. 8, but for simulations done with test charges instead of effective charges to compute electrostatic forces.

at the interprotein contact will become important. However, one important conclusion is that the interresidue contact distances for encounter definition obtained previously as 5.5 Å are equivalent to distances of 3.5–4 Å between interacting donor and acceptor atoms of these residues. That is, the same set of residue-residue positions are described by pairwise distances between several atoms of these residues that are shorter than the distances between specified atoms taken from the list in Table 2. This is further support for the proteins being in direct rather than solvent-mediated contact in the encounter complex.

Figs. 11 and 12 present the correlation between the experimental and computed rates for the mutants considered using the two-contact criterion. Although it results in computed rates that are approximately three times larger than the experimental values, the correlation is somewhat better on average at a contact distance of 6.25 Å ($r = 0.98$ for the log-log plot) than at a contact distance of 5.5 Å ($r = 0.94$). The *E60A* and *E73W* mutants approximately define a window of reasonable contact distances between 6.25 and 7.25 Å (see Fig. 7). Throughout this window, there are no more than two outliers in the correlation. At 6.25 Å they are *D54A* and *E73A*; and at 7.25 Å, they are *E60A* and the *E60A+D54A* double mutant. At 5.5 Å, all four of these mutants are outliers.

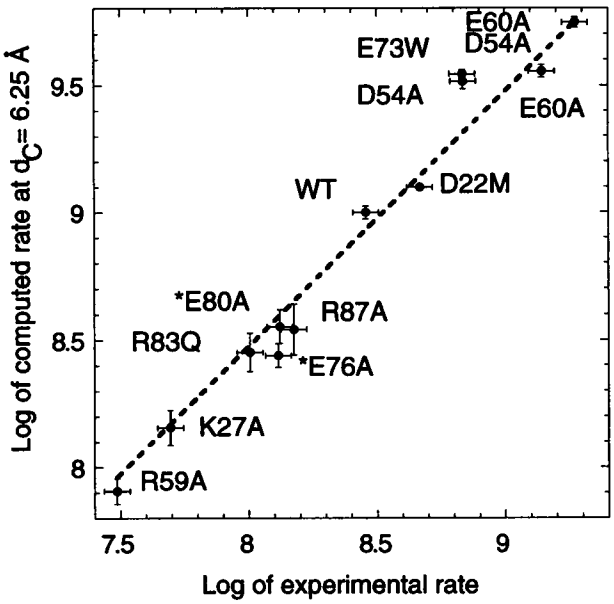

FIGURE 12 Computed versus experimental association rates multiplied by 3. Two contacts with the contact distance d_C = 6.25 Å are used to define the value of the computed rates.

DISCUSSION

Definition of the diffusional encounter complex

The process by which two proteins bind to each other can be considered as two consecutive steps. First, the proteins diffuse toward each other while being steered and oriented by long-range interactions. Then, when the proteins are close to each other, additional short-range interactions become important and the proteins adjust their positions to form a bound complex.

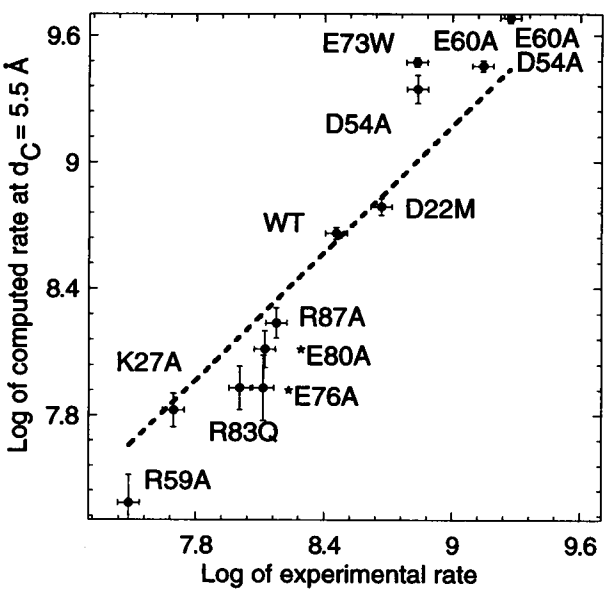

FIGURE 11 Computed versus experimental association rates for wild-type and 11 mutant proteins. Two contacts with the contact distance d_C = 5.5 Å are used to define the value of the computed rates.

In our simulations, we treat the proteins as rigid and account only for electrostatic and exclusion forces. Thus, while the first step can be modeled reasonably, the second cannot. While it is in principle possible to model the short-range interactions of the second step, and this has been done for simulations of enzyme-small substrate encounter (Luty et al., 1993; Luty and McCammon, 1993), this is not computationally feasible for realistic simulations of protein-protein encounter. This means that if experimental on-rates are to be reproduced by our simulations, the diffusional encounter complex must be formed during the first step. Once the diffusional encounter complex has been formed, subsequent rearrangements to form the bound state must occur effectively irreversibly, i.e., much faster than diffusional dissociation. Because the short-range interactions are not modeled fully, it is necessary to derive a criterion to determine when a diffusional encounter complex is formed during the simulations, i.e., when barnase and barstar are sufficiently closely aligned that they would automatically go on to form a bound complex when the necessary short-range interactions are present. The derivation of such a criterion is one of the main objects of this study.

One way to define the close alignment is by the RMS distance to the positions of atoms of the proteins in their bound complex. However, this definition would give incorrect results if, in the diffusional encounter complex, the two proteins are not in the exact vicinity of their positions in the bound complex. Furthermore, since the RMS distance cor-

responding to an encounter is rather large (6–7 Å) in the case of barnase and barstar, it is difficult to interpret in terms of physical intermolecular interactions.

The electrostatic interaction energy as computed here defines the encounter complex with about the same level of predictability as RMS distance. The simulations indicate that it gives a poor description of the diffusional encounter complex in the case of barnase and barstar. This would probably be true for other proteins because this criterion results in double counting of the effects of electrostatic steering.

The most successful approach in this work was to define diffusional encounter complex formation by the satisfaction of two correct residue-residue contacts. This definition can take into account drops in rates due to elimination (on mutation) of the contact which is formed most frequently. This is particularly important for the *R59A* mutation in barnase, for which the large decrease in rate cannot be accounted for by the decrease in the long-range steering forces alone. The *R59-E76* contact is formed most frequently during the simulations of the wild-type proteins. This provides an explanation for the large influence of the *R59A* mutation on the association rate. In contrast, mutation of residues *K27, R83*, or *R87* mostly changes the forces governing less frequently appearing contacts, and thus reduces the overall rate less than would be the case if encounter complex formation was exactly correlated with the position of the proteins in their bound complex.

Experimentally measured rates may be reproduced by assuming that two correct residue-residue contacts at 5.5 Å or three contacts at 7.0 Å are formed during diffusion. Structurally, there are no distinguishable differences between the encounter complexes obtained with two and three contacts at corresponding distances. However, the encounter defined by three contacts gives too large a contact distance for physical interpretation if this is derived by fitting experimental rates. Moreover, the rates for the *R59A* mutant appear to be overestimated.

Description of the diffusional encounter complex

Formation of the diffusional encounter complex for the wild-type proteins is mainly determined by only half of the eight contacts derived from the crystallographic complex coordinates. These are listed first in Table 2. These contacts are defined by the favorable electrostatic interaction region for the wild-type proteins seen in Fig. 2. The positions of the center of barstar in the encounter complexes generated during the simulations as defined with a two-contact criterion all fall within the uppermost yellow contour in Fig. 2 *A*. Since the simulated rates correlate with the experimental ones, which show a larger influence on the association rates for the *R59A* and *E60A* mutations than the analogous *K27A* and *D54A* mutations, this indicates that the guanine binding loop (residues 57–60) on barnase serves as a recognition site for barstar.

Computed rates

We were able to compute association rates within a factor of 2 of experimental values for the wt and 11 mutant proteins that have rates ranging over 2 orders of magnitude. The best correlation for the set of mutants was obtained when two contacts were required and the contact distance was increased to 6.25 Å. At this distance, the rates computed correlate with three times the experimental rates. It is reasonable to expect that computed diffusional rates should be somewhat higher than experimental rates because short-range interactions involved in formation of a bound complex are not modeled. The impact of such interactions was investigated by Northrup and Erickson (1992) for two model spherical proteins. From simulations, they found that if a short-range locking potential with a well depth of 4.2 kcal/mol was applied when two contacts were formed, a three-contact (for this system, bound) complex evolved as the trajectory continued in 50% of cases. The percentage of 33% for our simulations is similar.

From Fig. 7, it can be seen that most mutants have a similar dependence of the ratio of computed to experimental rates on the contact distance. The mutants having larger charges associate faster than expected if lower encounter distances are considered. At larger encounter distances, the situation is opposite, so that the intermediate distances (6–7 Å) suit all mutants equally well. However, the barnase mutant *E60A* is found to have only a slightly higher rate than mutants *D54A* and *E73W*, while the experimentally measured rate is ~2× larger. This systematic underestimation (although it is within the error estimate of modeling) is probably due to elimination of the contact formed by the residue *E60*, while during the association there might be contact interactions in that region even after mutation *E60A*. Simulations, where the contact is formally taken into account, indeed show that a factor of 2 increase in rate may be achieved by assuming the presence of the contact. On the other hand, mutation of either of the residues *D54* and *E73* is known to destabilize barnase by 2.5 kcal/mol (Meiering et al., 1992), i.e., significantly more than the other mutations we consider do. Therefore, possible conformational changes may lead to different interaction energies. These may result in reduced attraction between barstar and barnase and thus decrease the computed rates for these mutants.

The importance of the steering forces should be noted, since they are responsible for the increase in association rates of at least the factor of 20 modeled here on changing ionic strength from 500 to 50 mM. These forces are essentially long-ranged under 50-mM salt conditions. We can demonstrate this by carrying out simulations with the interactions artificially truncated. The rates are decreased fivefold when the forces acting on the charges on one of the proteins are truncated at distances >8 Å from the surface of the other protein. When the forces are truncated at distances <8 Å, one sees approximately the same fivefold decrease in the computed rate (Fig. 13). Consequently, the relative contribution of the forces at nearer and further distances is

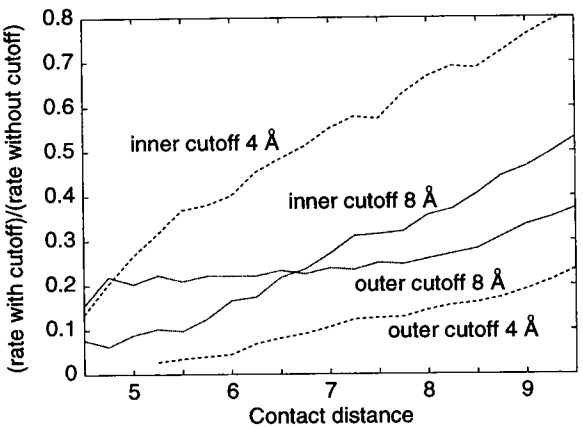

FIGURE 13 The rates computed for artificial systems, in which the interaction forces are cut at distances larger (*outer cutoff*) or smaller (*inner cutoff*) than the specified distances: 4 Å (*dots*) and 8 Å.

equal at ~8 Å, although both ranges of forces cause a fivefold enhancement of the rate compared to the rate computed in their absence. In contrast, the relative importance of forces from groups nearer or further than 4 Å is markedly different. The forces from groups beyond 4 Å contribute much more to electrostatic steering than the forces due to groups closer than 4 Å, although the latter are responsible for a threefold enhancement of rate.

Accuracy, sources of error, and model deficiencies in the simulations

Fig. 7 shows that, for 8 of 12 protein pairs, computed rates fit the experimental rates accurately independently of the adjustable parameter defining the encounter complex (the contact distance). Defining a contact distance of 6.25 Å, two more mutants (*E60A* and the *E60A* + *D54A* double mutant) are fit and the computed rates are 2–4× larger than the experimental rates (see the discussion above for why this overestimation is expected). Assuming that the association rate is given by $k_{on} \sim \exp(E/k_BT)$ (see Zhou, 1996) where E is the interaction energy in the encounter complex, this factor of 2 spread corresponds to an error of 0.7 k_BT. This error estimate is rather small if one takes into account the computed electrostatic interaction energies in the encounter complex of *10* k_BT, and that the electrostatic interaction energies must change with the conformations of the side chains. One may, however, appeal to gating theory (Northrup et al., 1982; Szabo et al., 1982) which states that only the optimal interactions (corresponding to an "open gate") should be in effect if the characteristic time for diffusional motion is much longer than the characteristic time for side-chain motions. It is also reasonable to assume that the x-ray coordinates of the bound complex provide protein conformations with optimal interactions. These can, therefore, be used to simulate association using a rigid model. The remaining surprise is that the same protein conformations are also suitable for the mutants modeled as crudely as we did.

The computed rates are modulated by parameters of the model to the same extent. For example, at 50 mM ionic strength, the treatment of the ionic solvent as defined by zero (instead of finite, 2.0 Å) ion radius decreases the rates by approximately a factor of 2. Moreover, we observe that two different protocols to model and optimize polar hydrogen positions may influence the computation results. The hydrogen atom positions may be optimized by energy minimization assuming that the proteins are in the crystallographically defined complex or that they are separated. The two sets of positions so derived have a small RMSD of 0.2 Å, but the rates computed using these two conformations differ by a factor of 2.

Treating flexibility explicitly would cause two major differences in the model. First, the encounter would be modeled more realistically as a set of dynamic contacts. Second, the conformational changes would result, at least for wild-type proteins, in protein interactions that are less optimal for association. The contact distances required to reproduce experimental rates with rigid protein models are quite large, 5–6 Å, making the computation results rather insensitive to the positions of the side chains used to define the reaction. For example, altering the conformation of one residue, *R59* in barnase, did not cause significant changes in the computed rates. Test calculations with coordinates of the unbound barnase from x-ray crystallography (Buckle et al., 1993, pdb identifier 1bnj) and NMR studies (Bycroft et al., 1991, pdb structure 1bnr), showed, however, that both interactions and reaction criteria may highly disfavor association when the conformations of the two proteins are not complementary. The rates calculated with the different members of the ensemble of structures varied widely, yielding up to a factor of 10 decrease in the computed rates. Calculations with complementary conformations obtained from short (40 ps) molecular dynamics simulations of the barnase-barstar complex resulted in approximately 2× smaller rates than those computed with the x-ray structure. They did not, however, in these preliminary calculations, change the correlation between experimental and computed rates for the wild-type proteins and *E73W* and *D54A* mutants (which were the only mutants investigated).

We used the linearized Poisson-Boltzmann description of ionic salt effects. The use of a nonlinear description gave similar results, indicating that the presence of monovalent salt (NaCl in experiments) can be modeled with a linear treatment to within the error limits of the present computations (see also Schreiber and Fersht, 1996).

Hydrodynamic interactions (Brune and Kim, 1994) between proteins were not accounted for. For model systems mimicking protein-protein association, they have been estimated to alter association rates by up to ~20% (Antosiewicz and McCammon, 1995; Antosiewicz et al., 1996). We expect their effect on association rates to be smaller than the factor-of-2 errors in our computed rates and to have

a negligible effect on the relative rates of the different mutants.

CONCLUSIONS

In order to model the diffusional association of proteins by Brownian dynamics simulation successfully, it is important to define detailed and appropriate criteria for interprotein encounter complex formation. In this study, the most successful criterion for encounter complex formation is based on satisfaction of residue-residue contacts present in the target complex. The structure of a target complex is required to define this criterion.

The requirement for the formation of two or three correct interresidue contacts is found to be appropriate, i.e., the diffusion rates for two and three contacts correlate well with experimental data for the majority of the mutant proteins studied. The physical basis for these contacts is hydrogen-bonding between donor and acceptor atoms in the two proteins. Thus, this type of criterion can also be derived for other proteins.

With this criterion, the positions of barnase and barstar in the encounter complex are close to their crystallographic positions, but barstar tends to be shifted toward the guanine binding loop. This loop appears to be the most important local region of electrostatic attraction during diffusion. Long-range electrostatic steering makes a large contribution to the diffusional association rates in this system. It is the explanation for the 20-fold increase in association rates on changing ionic strength from 500 to 50 mM.

We are grateful to A. R. Fersht and G. Schreiber for making experimental data available before publication, as well as for useful discussions. We thank B. Das, G. Schreiber, and D. Walther for critical reading of the manuscript. The contribution of D. Starks-Browning in optimization of the Brownian dynamics code for the SGI Power Challenge is gratefully acknowledged.

REFERENCES

Allison, S. A., S. H. Northrup, and J. A. McCammon. 1986. Simulation of biomolecular diffusion and complex formation. *Biophys. J.* 49:167–175.

Antosiewicz, J., J. M. Briggs, and J. A. McCammon. 1996. Orientational steering in enzyme-substrate association: ionic strength dependence of hydrodynamic torque effects. *Eur. Biophys. J.* 24:137–141.

Antosiewicz, J., and J. A. McCammon. 1995. Electrostatic and hydrodynamic orientational steering effects in enzyme-substrate association. *Biophys. J.* 69:57–65.

Berg, O. G., and P. H. von Hippel. 1985. Diffusion-controlled macromolecular interactions. *Ann. Rev. Biophys. Biophys. Chem.* 14:131–160.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data base: a computer-based archival file for macromolecular structure. *J. Mol. Biol.* 112:535–542.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Brune, D., and S. Kim. 1994. Hydrodynamic steering effects in protein association. *Proc. Natl. Acad. Sci. U.S.A.* 91:2930–2934.

Buckle, A. M., K. Henrick, and A. R. Fersht. 1993. Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* 234:847.

Buckle, A. M., G. Schreiber, and A. R. Fersht. 1994. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0 Å resolution. *Biochemistry.* 33:8878–8889.

Bycroft, M., S. Ludvingen, A. R. Fersht, and F. M. Poulsen. 1991. Determination of the solution conformation of barnase using NMR spectroscopy. *Biochemistry.* 30:8697–8701.

Calef, D. F. 1983. Diffusion-controlled reactions. *Ann. Rev. Phys. Chem.* 34:493–524.

Creighton, T. E. 1993. Protein: structures and molecular properties. W. H. Freeman and Co., New York.

Davis, M. E., and J. A. McCammon. 1989. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comp. Chem.* 10:386–391.

Davis, M. E., and J. A. McCammon. 1990. Calculating electrostatic forces from grid-calculated potentials. *J. Comp. Chem.* 11:401–409.

DeLisi, C. 1980. The biophysics of ligand-receptor interactions. *Q. Rev. Biophys.* 13:201–230.

Ermak, D. L., and J. A. McCammon. 1978. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* 69:1352–1360.

Gabdoulline, R. R., and R. C. Wade. 1996. Effective charges for macromolecules in solvent. *J. Phys. Chem.* 100:3868–3878.

Gosting, L. J. 1956. Measurement and interpretation of diffusion coefficients of proteins. *Adv. Protein Chem.* 11:429–554.

Hartley, R. W. 1993. Directed mutagenesis and barnase-barstar recognition. *Biochemistry.* 32:5978–5984.

Ito, W., H. Yasui, and Y. Kurosawa. 1995. Mutations in the complimentarity determining regions do not cause differences in free energy during the process of formation of the activated complex between the antibody and the corresponding protein antigen. *J. Mol. Biol.* 248:729–732.

Johnstone, R. W., S. M. Andrew, M. P. Hogarth, G. A. Pietersz, and I. F. C. McKenzie. 1990. The effect of temperature on the binding kinetics and equilibrium constants of monoclonal antibodies to cell surface antigens. *Mol. Immunol.* 27:327–333.

Jorgensen, W., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. Energy minimization for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

Jucovic, M., and R. W. Hartley. 1996. Protein-protein interaction: A genetic selection for compensating mutations at the barnase-barstar interface. *Proc. Natl. Acad. Sci. U.S.A.* 93:2343–2347.

Kozack, R. E., M. J. d'Mello, and S. Subramaniam. 1995. Computer modeling of electrostatic steering and orientational effects in antibody-antigen association. *Biophys. J.* 68:807–814.

Kozack, R. E., and S. Subramaniam. 1993. Brownian dynamics simulations of molecular recognition in an antibody-antigen system. *Protein Sci.* 2:915–926.

Kraulis, P. 1991. MolScript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.

Luty, B. A., E. Amrani, and J. A. McCammon. 1993. Simulations of the bimolecular reaction between superoxide and superoxide dismutase: synthesis of the encounter and reaction steps. *J. Am. Chem. Soc.* 115: 11874–11877.

Luty, B. A., and J. A. McCammon. 1993. Simulations of bimolecular reactions: synthesis of the encounter and reaction steps. *Molecular Simulations.* 10:61–65.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatic and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comp. Phys. Comm.* 91:57–95.

Madura, J. D., M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon. 1994. Biological applications of electrostatic calculations and Brownian dynamics simulations. *Rev. Comp. Chem.* V:229–267.

Malmqvist, M. 1993. Surface plasmon resonance for detection and measurement of antibody-antigen affinity and kinetics. *Curr. Opin. Immunol.* 5:282–286.

McCammon, J. A., S. H. Northrup, and S. A. Allison. 1986. Diffusional dynamics of ligand-receptor association. *J. Phys. Chem.* 90:3901–3905.

Meiering, E. M., L. Serrano, and A. R. Fersht. 1992. Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* 225:585–589.

Nambi, P., A. Wierzbicki, and S. A. Allison. 1991. Intermolecular interaction between bovine pancreatic trypsin inhibitor molecules probed by Brownian dynamics simulation. *J. Phys. Chem.* 95:9595–9600.

Northrup, S. H. 1994. Hydrodynamic motions of large molecules. *Curr. Opin. Struct. Biol.* 4:269–274.

Northrup, S. H., S. A. Allison, and A. A. McCammon. 1983. Brownian dynamics simulations of diffusion influenced bimolecular reactions. *J. Chem. Phys.* 80:1517–1524.

Northrup, S. H., J. O. Boles, and J. C. L. Reynolds. 1987. Electrostatic effects in the Brownian dynamics of association and orientation of heme proteins. *J. Phys. Chem.* 91:5991–5998.

Northrup, S. H., J. O. Boles, and J. C. L. Reynolds. 1988. Brownian dynamics of cytochrome *c* and cytochrome *c* peroxidase association. *Science.* 241:67–70.

Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 89:3338–3342.

Northrup, S. H., K. A. Thomasson, C. M. Miller, P. D. Barker, L. D. Eltis, J. G. Guillemette, S. C. Inglis, and A. G. Mauk. 1993. Effects of charged amino-acid mutations on the bimolecular kinetics of reduction of yeast iso-1-ferricytochrome *c* by bovine ferrocytochrome b_5. *Biochemistry.* 32:6613–6623.

Northrup, S. H., F. Zarin, and J. A. McCammon. 1982. Rate theory for gated diffusion-influenced ligand binding to proteins. *J. Phys. Chem.* 86:2314–2321.

Pollard, J. H. 1976. A handbook of numerical and statistical techniques. Cambridge University Press, Cambridge, UK.

QUANTA molecular modeling software package. 1992. Molecular Simulations, Inc., San Diego, CA.

Schreiber, G., and A. R. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry.* 32:5145–5150.

Schreiber, G., and A. R. Fersht. 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* 248:478–486.

Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3:427–431.

Smoluchowski, M. V. 1917. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Loeschungen. *Z. Phys. Chem.* 92: 129–168.

Stone, S. R., S. Dennis, and J. Hofsteenge. 1986. Quantitative evaluation of the contribution of ionic interactions to the formation of the thrombin-hirudin complex. *Biochemistry.* 28:6857–6863.

Szabo, A., D. Shoup, S. H. Northrup, and J. A. McCammon. 1982. Stochastically gated diffusion-influenced reactions. *J. Chem. Phys.* 77: 4484–4493.

Szabo, A., L. Stolz, and R. Granzow. 1995. Surface plasmon resonance and its use in bimolecular interaction analysis (BIA). *Curr. Opin. Struct. Biol.* 5:699–705.

Wade, R. C. 1996. Brownian dynamics simulations of enzyme-substrate encounter. *Biochem. Soc. Trans.* 24:254–259.

Wade, R. C., B. A. Luty, E. D. Demchuk, J. D. Madura, M. E. Davis, J. M. Briggs, and J. A. McCammon. 1994. Simulation of enzyme-substrate encounter with gated active sites. *Nature Struct. Biol.* 1:65–69.

Ward, E. S., D. Güssow, A. D. Griffiths, P. T. Jones, and G. Winter. 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli. Nature.* 341:544–546.

Zhou, H.-X. 1990. On the calculation of diffusive reaction rates using Brownian dynamics simulations. *J. Phys. Chem.* 92:3092–3095.

Zhou, H.-X. 1993. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* 64:1711–1726.

Zhou, H.-X. 1996. Effect of interaction potentials in diffusion-influenced reactions with small reactive regions. *J. Chem. Phys.* 105:7235–7237.