METHODS: A Companion to Methods in Enzymology **14**, 329–341 (1998)
Article No. ME980588

Brownian Dynamics Simulation of Protein–Protein Diffusional Encounter

Razif R. Gabdoulline and Rebecca C. Wade[1]

Structural Biology Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Protein association events are ubiquitous in biological systems. Some protein associations and subsequent responses are diffusion controlled *in vivo.* Hence, it is important to be able to compute bimolecular diffusional association rates for proteins. The Brownian dynamics simulation methodology may be used to simulate protein–protein encounter, compute association rates, and examine their dependence on protein mutation and the nature of the physical environment (e.g., as a function of ionic strength or viscosity). Here, the theory for Brownian dynamics simulations is described, and important methodological aspects, particularly pertaining to the correct modeling of electrostatic forces and definition of encounter complex formation, are highlighted. To illustrate application of the method, simulations of the diffusional encounter of the extracellular ribonuclease, barnase, and its intracellular inhibitor, barstar, are described. This shows how experimental rates for a series of mutants and the dependence of rates on ionic strength can be reproduced well by Brownian dynamics simulations. Potential future uses of the Brownian dynamics method for investigating protein–protein association are discussed. © 1998 Academic Press

Protein–protein association is fundamental to processes such as signal transduction, transcription, cell cycle regulation, and immune response. The speed of the association phase puts an upper limit on the speed of the response resulting from protein binding *in vivo.* Frequently, at least one of the interacting proteins is free to move in the intra- or extracellular environment and must find its binding partner by diffusion. The association rate is limited by the time required to bring the "reactive patches" of the proteins together by diffusion (1, 2) so that the proteins form an "encounter complex." When the postdiffusional step of the association process is much faster than diffusional dissociation, the "reaction" is diffusion controlled; that is, the association rate of the two proteins is defined by their diffusional encounter (3). The determination of whether a reaction or response is diffusion controlled is clearly crucial to protein design studies.

Recently, there has been quite a dramatic increase in the number of experimental measurements of the kinetics of protein–protein interactions (4). This is in part due to the recent availability of instruments based on surface plasmon resonance (4, 5), which, in principle, provide a generally applicable noninvasive method to measure the kinetics of protein–protein interactions without the necessity for labeling. Experimentally measured association rates cover a wide range from 10^3 to 10^9 M^{-1} s^{-1}. The lower limit is dictated largely by experimental limits on the rates measurable (4), while the upper limit is based on measured association rates in solution for the association of thrombin and hirudin (6) and of barnase and barstar (7). The following properties of protein association rates indicate diffusion control:

a. Fast, i.e., $\geq 10^6$ M^{-1} s^{-1} under typical conditions.
b. Inverse dependence on solvent viscosity (1) and, therefore, linear dependence on the proteins' relative diffusion constant.
c. Dependence on the ionic strength of the solution, indicating the importance of long-range electrostatic forces.
d. Temperature dependence following the temperature dependence of the solvent viscosity. The diffusion constants of proteins in water approximately double over the temperature range 288–313 K (8) , and, as has been observed for some antibody–antigen complexes (9, 10), this causes diffusion-controlled rates to double too.
e. Sensitivity to the diffusional environment, e.g., whether, before complexation, both proteins are free to diffuse in solution or one of them is immobilized on a surface.

None of these five properties alone or in combination is a proof of diffusional control. Therefore, detailed sim-

[1] To whom correspondence should be addressed. Fax: +49 6221 387 517. E-mail: wade@embl-heidelberg.de.

1046-2023/98 $25.00

ulation of diffusional association is a necessary step in defining the mechanism of interaction. Brownian dynamics (BD) simulations provide a means to compute bimolecular diffusional association rates . They have been used for calculating protein–protein and enzyme–substrate association rates using both atomic-detail models and highly simplified models [for reviews, see (11, 12)]. BD simulations provide insights into the properties and mechanisms of diffusion-controlled association and have been used for the design of protein mutants with altered association rates [see, e.g., (13)].

BD simulations by Northrup and Erickson (14) of the association of model spherical proteins showed that associations rates in the absence of long-range attractive forces should be $\sim 2 \times 10^6$ M^{-1} s^{-1}. This is about 1000-fold faster than would be expected solely on the basis of geometric criteria for bringing the "reactive patches" on the proteins into contact. However, in a diffusional system the particles do not rebound away from each other after colliding with a nonreactive contact as in a Newtonian system; instead, they diffuse close to each other, making multiple collisions and rotationally reorienting between them. This diffusive entrapment effect results in fast association rates of the order 10^6 M^{-1} s^{-1}, which are only about 1000-fold slower than the theoretical diffusion-limited Smoluchowski rate for two isotropically reactive spheres. Rates of $\sim 10^6$ M^{-1} s^{-1} are often seen for protein–protein complexation, e.g., antibody–antigen association. However, rates exceeding 10^6 M^{-1} s^{-1} and ranging up to $\sim 10^9$ M^{-1} s^{-1} are observed for some protein–protein encounters, indicating that electrostatic interactions facilitate binding in these cases. Electrostatic steering is the main focus of BD simulations of protein–protein association.

In this paper, we first present the theory for simulating protein diffusion by BD and computing bimolecular association rates. We then describe how simulations are performed, highlighting the most important technical aspects. We illustrate this with a case study of the association of barnase and barstar. We conclude by outlining other systems to which either the method has been applied or that provide possible subjects of future study.

THEORY

Simulation of Diffusional Motion by Brownian Dynamics

The theory of Brownian motion describes the dynamic behavior of particles, the mass and size of which are larger than those of the molecules of the solvent in which they are immersed. These particles are subject to stochastic collisions with the solvent molecules which lead to the apparently random motion of the particles that is diffusion. Such motion was first recorded in 1827 by Brown, who observed erratic motions of pollen grains in water. Einstein (15) and von Smoluchowski (16) showed that the displacement $\Delta\mathbf{r}$ of a particle undergoing Brownian motion in time Δt is given by

$$\langle \Delta\mathbf{r}^2 \rangle = 6D\Delta t, \qquad [1]$$

where D is the translational diffusion coefficient of the particle and

$$D = k_b T / 6\pi\eta a, \qquad [2]$$

where k_b is Boltzmann constant, T is absolute temperature, η is solvent viscosity, and a is the hydrodynamic radius of the particle.

The dynamics of diffusional motion are described by the Langevin equation and there are a number of ways to solve this equation (17). In simulations of protein–protein association, the method used is that presented by Ermak and McCammon in 1978 for simulation of the motion of Brownian particles (18). In this method, a trajectory is generated as a set of snapshots of the particles at time intervals of Δt. The positions of the particles are computed from the Ermak–McCammon equations for translational and rotational motion. In their simplest forms, these are as follows:

- The translational displacement $\Delta\mathbf{r}$ of each particle during each time step is given by

$$\Delta\mathbf{r} = (k_b T)^{-1} D\mathbf{F}\Delta t + \mathbf{R}, \qquad [3]$$

where $\mathbf{F}$ is the systematic force on the particle before the step is taken, and $\mathbf{R}$ is a random vector satisfying $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R}^2 \rangle = 6D\Delta t$. The $(k_b T)^{-1} D$ factor by which $\mathbf{F}$ is multiplied models the damping effect of solvent friction.

- The rotational displacement angle $\Delta\mathbf{w}$ of each particle at each time step is given by

$$\Delta\mathbf{w} = (k_b T)^{-1} D_R \mathbf{T}\Delta t + \mathbf{W}, \qquad [4]$$

where $\mathbf{T}$ is the torque acting on the particle before the step is taken, D_R is the rotational diffusion constant of the particle, and $\mathbf{W}$ is a random rotation angle satisfying $\langle \mathbf{W} \rangle = 0$ and $\langle \mathbf{W}^2 \rangle = 6D_R\Delta t$.

For simulations of protein–protein interactions in which the proteins are treated as rigid bodies, there are only two solute particles. Therefore, translational motion is simulated for one of the proteins (protein II) relative to the position of the other (protein I) (see Fig.

1). The displacement of protein II is given by Eq. [3], with D replaced by the relative translational diffusion constant.

The effects of hydrodynamic interactions are usually neglected for simulations of protein–protein association but can be treated by substituting tensors for the diffusion constants in the above equations.

To use these equations to propagate diffusional motion, the time step, Δt, must be small enough that the forces and torques on the particles (and the gradient of the diffusion tensors) remain effectively constant during the time step. On the other hand, the motion can be analyzed only over periods exceeding a particle's momentum relaxation time, i.e., $\Delta t \gg mD/k_bT$, where m is the mass of the particle.

Computation of Bimolecular Diffusional Association Rates

Computation of bimolecular association rates requires solution of the diffusion equation. This can be solved analytically only for systems with simple geometry and charge distribution. Numerical BD simulations are necessary to permit evaluation of the effects on rate constants of a protein's complex shape, heterogeneous charge distribution, internal motions, hydrodynamic interactions, and nonuniform reactivity. Nevertheless, "boundary conditions" given by an analytical solution to the diffusion equation are necessary to use the results of BD simulations to compute rates. Thus, the association rate k computed from a BD simulation is given by

$$k = k_D(b)\beta^\infty, \qquad [5]$$

where $k_D(b)$ is the steady-state rate constant for two particles approaching within a distance b (see Fig. 1) and β^∞ is the probability that having reached this separation, the particles will go on to "react" and form an encounter complex rather than diffuse apart to infinite separation. $k_D(b)$ must be computed analytically, whereas β^∞ is computed from BD simulations in which thousands of trajectories are generated, each of which is started with the particles separated by distance b and is truncated when they reach a separation distance q.

For two spherical particles with relative diffusion constant D and no interparticle forces, $k_D(b)$ is given by the Smoluchowski expression (19):

$$k_D(b) = 4\pi Db. \qquad [6]$$

When interparticle forces are present,

$$k_D(b) = 4\pi\left[\int_b^\infty \left(\frac{e^{\{E(r)/k_BT\}}}{r^2D}\right)dr\right]^{-1}, \qquad [7]$$

where $E(r)$ is the centrosymmetric interaction energy between the particles, which depends on their separation distance r.

β^∞ is usually given by (20)

$$\beta^\infty = \beta[1 - (1 - \beta)\Omega], \qquad [8]$$

where $\Omega = k_D(b)/k_D(q)$ and β is the fraction of trajectories in which encounter complex formation occurs before the particles diffuse to a separation distance of q. The multiplier for β in Eq. [8] is a correction factor to account for the truncation of trajectories at a finite separation distance, q.

Several alternative ways to compute association rates using BD simulations have been developed to improve efficiency, e.g., the WEBDS (Weighted-Ensemble Brownian Dynamics Simulations) method (21), an algorithm to compute time-dependent rate coefficients (22), and an algorithm to remove the outer sphere by using an "m surface" (23). These methods have been applied either to simulation of simple model systems for protein–protein association (24) or to atomic-detail simulations of enzyme–substrate association (25). They

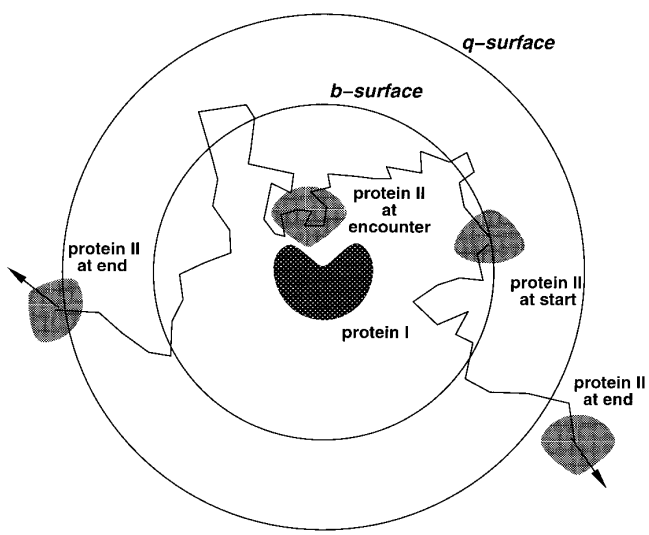

FIG. 1. Schematic diagram indicating the setup of the system for BD simulations to compute bimolecular diffusion-controlled rate constants. Simulations are conducted in coordinates defined relative to the position of the central protein (protein I). At the beginning of each trajectory, the second protein is positioned with a randomly chosen orientation at a randomly chosen point on the surface of the inner sphere of radius b. BD simulation is then performed until this protein diffuses outside the outer sphere of radius q. During the simulation, satisfaction of reaction criteria for encounter complex formation is monitored. The radius b is chosen so that outside the sphere the forces between the proteins are centrosymmetric and inside they are anisotropic. The radius q is chosen so that the inward reactive flux at separation q is centrosymmetric (22).

have not yet been applied to simulations of protein–protein association using detailed protein models.

Modeling of Forces Governing Diffusional Motion of Proteins

For simulations of protein–ligand association, intermolecular forces and torques are given by the sum of electrostatic and exclusion forces. Short-range attractive interactions such as hydrogen bonding and van der Waals interactions are not modeled because they are not as important for encounter complex formation as for formation of the final bound complex and because it is too computationally demanding to model them over the time scales of BD simulations. These time scales are orders of magnitude longer than those for molecular dynamics simulations in which a more complete model of forces is used. Additional forces that are sometimes modeled for BD simulation of bimolecular associations involving proteins are hydrodynamic interactions and intramolecular forces to model internal flexibility.

Electrostatic Forces

A range of electrostatic models have been employed in BD simulations of protein–ligand association. A continuum model of the solvent is adopted for the BD simulations and it is also used for computing electrostatic forces. The electrostatic models, however, vary in the level of the detail with which the charge distribution in the proteins is modeled and the approximations used to represent the dielectric environment of the system.

In simple models, used mostly in early simulations (26), a small number of charges are positioned in the proteins so as to reproduce their monopole, dipole, and quadrupole moments, and interactions between charges are computed using Coulomb's law either with a constant dielectric permittivity or with a distance-dependent dielectric permittivity to implicitly model the dielectric variation in the system simulated. Ionic strength was modeled by Debye–Hückel screening.

More recently, it has become possible to assign partial atomic charges to all atoms in the proteins and to model the dielectric heterogeneity and ionic strength explicitly by solving the Poisson–Boltzmann equation numerically:

$$\nabla(\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})) - c(\mathbf{r})\epsilon(\mathbf{r})\kappa^2\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}). \qquad [9]$$

Here, the dielectric constant, $\epsilon(\mathbf{r})$, is a function of position, $\phi(\mathbf{r})$ is the electrostatic potential, $\rho(\mathbf{r})$ is the charge density, κ is the inverse of the Debye–Hückel screening length (11), and $c(\mathbf{r})$ defines the volume accessible to counterions; it is 0 inside the solute and the Stern layer and 1 outside. Aqueous solvent is assigned a high relative dielectric constant of about 80, whereas the proteins are assigned lower dielectric constants, typically 2–4. The boundary between these dielectrics is assigned from atomic coordinates and radii. While there are a number of numerical methods for solving the Poisson–Boltzmann equation, the most commonly used is the finite-difference method in which a set of difference equations of the following type are solved iteratively (27):

$$h^2 \sum \frac{\epsilon_{ij}(\phi_i - \phi_j)}{h} + c_i h^3 \epsilon_i \kappa^2 \phi_i = 4\pi q_i. \qquad [10]$$

Here, j runs over the six grid points adjacent to grid point i, ϵ_{ij} is the permittivity of the face connecting i and j, c_i is the fraction of the volume accessible to counterions, ϵ_i is the permittivity at point i, and q_i is the charge enclosed. The parameters of this grid will influence the accuracy of the calculation (see below).

The finite-difference method can be used to compute the intermolecular free energy, and hence the intermolecular forces, for a system consisting of two low dielectric solutes in a high-dielectric solvent. This is, however, too computationally demanding for use in BD simulations for which forces are required at every time step in thousands of trajectories. Instead, the test charge approximation has usually been adopted. For this, the electrostatic potential of one protein (protein I) is computed on a grid. The second protein (protein II) then moves on the potential grid of protein I and forces are computed considering protein II as a collection of point charges in the solvent dielectric; i.e., the low dielectric and ion exclusion volume of protein II are ignored. We have recently shown (28) that forces in BD simulations can be computed more accurately but with virtually the same computational cost as the test charge approximation by using effective potential-derived charges instead of test charges. To derive "effective charges," 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, taking into account the inhomogeneous dielectric medium and the surrounding ionic solvent. Then effective charges are computed for each protein by fitting so that they reproduce its external potential in a homogeneous dielectric environment corresponding to that of the solvent. To compute the forces and torques acting on protein II (I), the array of effective charges for protein II (I) is placed on the electrostatic potential grid of protein I (II). This procedure 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 the diameter of the solvent probe and the desolvation energies of the charges in one protein due to the low dielectric cavity of the other protein become significant (28).

Exclusion Forces

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

Hydrodynamic Interactions

Hydrodynamic interactions (HIs) between particles in a solvent arise because of solvent flow induced by the motions of the particles. They can be incorporated into BD simulations by using a diffusion tensor [e.g., Oseen or Rotne-Prager (29)] but their computational requirements increase rapidly with the number of hydrodynamic centers. Their magnitude and importance for protein–ligand association have been studied for simple protein models. The estimated effects on rates range from 15% (30–33) to 50% (34) to two-orders of magnitude (35), and depend on the treatment of hydrodynamic boundary conditions and the sizes, shapes, and velocities of the solute molecules for which HI effects were computed. The smaller value appears to hold for conditions relevant for protein–protein association. HI can both increase and decrease molecular association rates. For example, consider an elongated protein binding in a slitlike cleft of another protein. HI will tend to lower the rate of translational diffusion toward the binding site. However, HI torques will favor binding of the elongated protein when it is nearly aligned with the cleft by enhancing the alignment. Thus, the effects of different HIs on association rates may tend to cancel each other out.

Modeling Flexibility

In simulations of protein–protein association performed to date, the proteins have been treated as rigid bodies. The effects of flexibility have been considered only by generating different sets of trajectories using different protein conformations [see (36) and below].

For simulations of enzyme–substrate encounter, other representations of protein flexibility have been used and these could be applied to simulations of protein–protein encounter:

—Reduction of atomic radii at the binding site to implicitly mimic the flexibility of these atoms (25).

—Modeling the motion of particularly flexible regions, such as protein loops, explicitly during the BD simulations. A simplified model is necessary to make this computationally feasible. A model treating a peptide loop as a chain of spheres, each sphere representing an amino acid residue, has been used (37). The spheres interact via a set of forces designed to mimic the geometry and interactions of an all-atom protein model.

METHOD

Procedure for Setting Up and Running BD Trajectories

The procedure for carrying out a BD simulation to compute a bimolecular association rate is outlined in Fig. 2. The main steps follow.

a. Model Protein Coordinates

If starting from a crystal structure, it will be necessary to add polar hydrogen atoms to the structure and possibly model in residues that were not defined in the electron density. Protons should be correctly positioned for the relevant pH and their assignment may be helped by computing the pK_a values of the titratable groups in the protein [see (38) and references therein]. The proton positions should be optimized by energy minimization, and perhaps a Monte-Carlo (39) or molecular dynamics optimization. As mentioned below, even small adjustments in their positions can alter rates (36).

b. Assign Atomic Parameters

The protein atoms should be assigned charges and radii. These can come either from molecular mechanics force fields [e.g., OPLS (40), CHARMM (41)] or from parameter sets derived specifically for use with a continuum solvent model [PARSE (42)].

c. Compute Electrostatic Potentials for Both Molecules

Solvent and protein dielectric constants, ionic strength, width of the ion exclusion (Stern) layer, and temperature of the system should be assigned. Potentials should be computed on grids large enough to include the volume over which the potential is nonisotropic. If this is very large, an additional inner grid with smaller spacing should be computed and both grids should be used in the BD simulations. For efficiency of handling, grid size should not exceed the size limited by computer memory requirements; for accuracy, the spacing for the inner grid should not exceed 1 Å (and smaller grid spacings are preferable).

d. Compute Effective Charges

Effective charge sites are usually assigned to the positions of the titratable residues (thereby reducing the number of charge sites), and their magnitude and sign are determined by fitting to reproduce the electrostatic potential in a shell around the protein using a homogeneous dielectric.

e. Choose b and q Surfaces

The b surface should be chosen by examining the electrostatic potentials so that at a distance b from each protein, the electrostatic forces on the other protein are

isotropic and centrosymmetric. The *q* surface should be placed much further away where the diffusive flux of the molecules is centrosymmetric (see Fig. 1). Typical distances are 50–100 Å for the *b* surface and 100–500 Å for the *q* surface. Sensitivity to alteration of these values should be tested during subsequent BD simulations.

f. Compute Exclusion Volumes and Surface Atoms

The exclusion volume is computed on a grid for one protein (protein I) by assigning values of 1 to grid points inside a probe-accessible surface and 0 outside. A spacing of 1 Å is usually adequate, but a 0.5-Å spacing may give better convergence and smoother, shorter trajectories. The surface-exposed atoms of the other protein (protein II) are listed. Steric overlap is defined to occur when one of the surface-exposed atoms is projected onto a grid point with value 1 (43). This computation is much quicker than a pairwise evaluation of exclusion forces, but requires the assumption that all atoms of protein II have the same radius as the probe used to define the probe-accessible surface of protein I.

g. Define Reaction Criteria

The choice of reaction criteria has a large effect on rates computed and they must therefore be chosen with care. Reaction criteria are required to define formation of a diffusional encounter complex. This is the complex

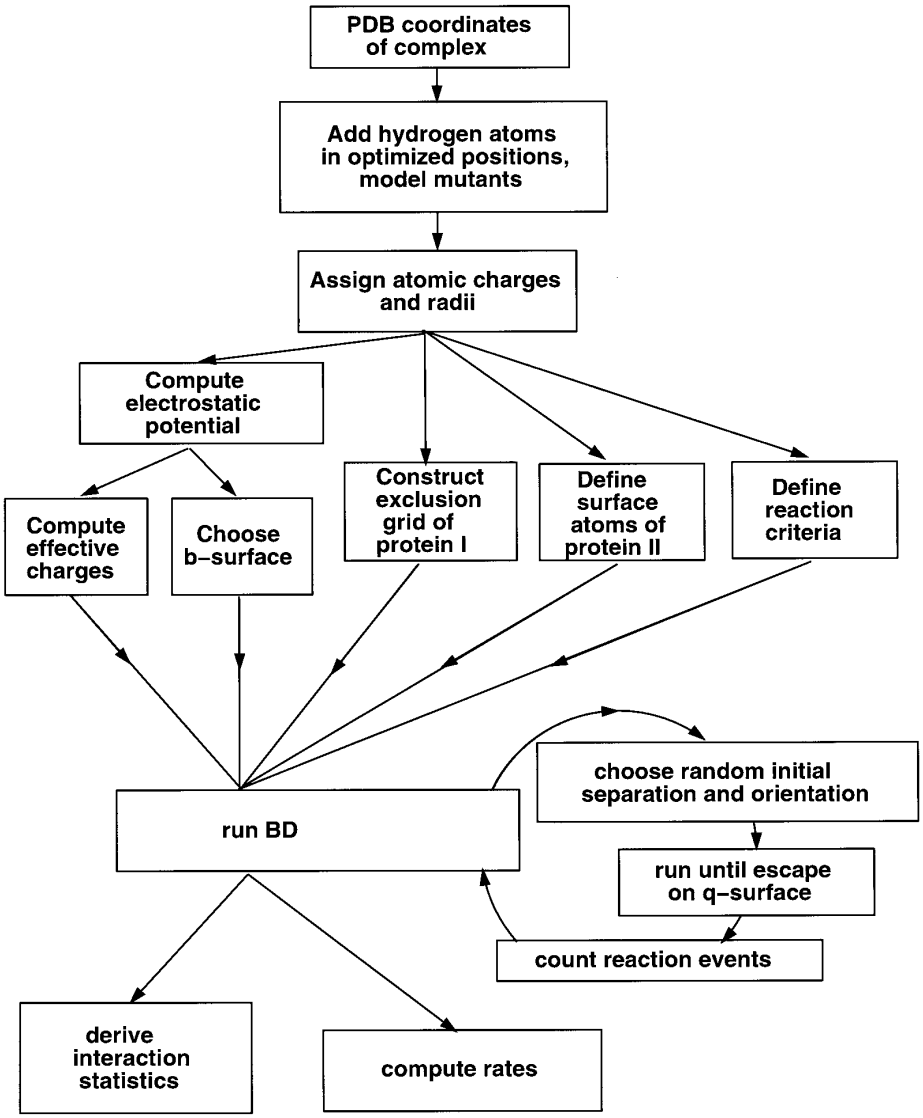

FIG. 2. Schematic diagram showing the steps in the computation of bimolecular diffusion-controlled rate constants by Brownian dynamics simulation.

formed at the endpoint of the diffusional association phase from which, in the next phase of binding, the proteins would rearrange into a tightly bound complex. Complete binding does not occur in the BD simulations because of the incompleteness of the force model used. One or several reaction criteria may be monitored during a trajectory. The criteria may be energetic (e.g., based on electrostatic energy) or geometric (rmsd of atoms from specified position, arrangement between protein atoms or cofactors) or based on contacts (number of contacts, buried surface area) (see Fig. 3). A criterion based on the arrangement of heme cofactors has been used for electron transfer proteins (44, 45). More generally, we have found that the energetic criterion should not be used and that a criterion based on formation of two to three correct contacts performs best for protein–protein association (see below for details).

h. Define BD Parameters

Rotational and translational diffusion constants should be defined. This can be done most simply using the Stokes–Einstein relationship and treating each protein as a hydrodynamically isotropic sphere so that its diffusion constants are defined by sphere radius, solvent viscosity, and temperature. More sophisticated estimations of diffusion constants account for protein surface shape (46, 47). A variable time step should be used that is smaller when the proteins are closer to each other and when protein II is close to the *q* surface. The smallest time step for protein–protein simulations with rigid body proteins should typically be 0.5–1.0 ps. For efficiency, rotations may be performed less frequently than translations. The number of trajectories run will depend on the accuracy required and the magnitude of the rates computed. The slower the rates, the greater the number of trajectories that will be required. Statistical errors in rates can be estimated either from Poisson statistics as $K^{-1/2} \cdot 100\%$, where K is the number of encounter events, or by subdividing the trajectories, computing rates for each subdivision, and then deriving the error as the rmsd of the rates from their average.

i. Run BD Simulations

Each simulation is started with a randomly chosen position and orientation of protein II on the *b* surface

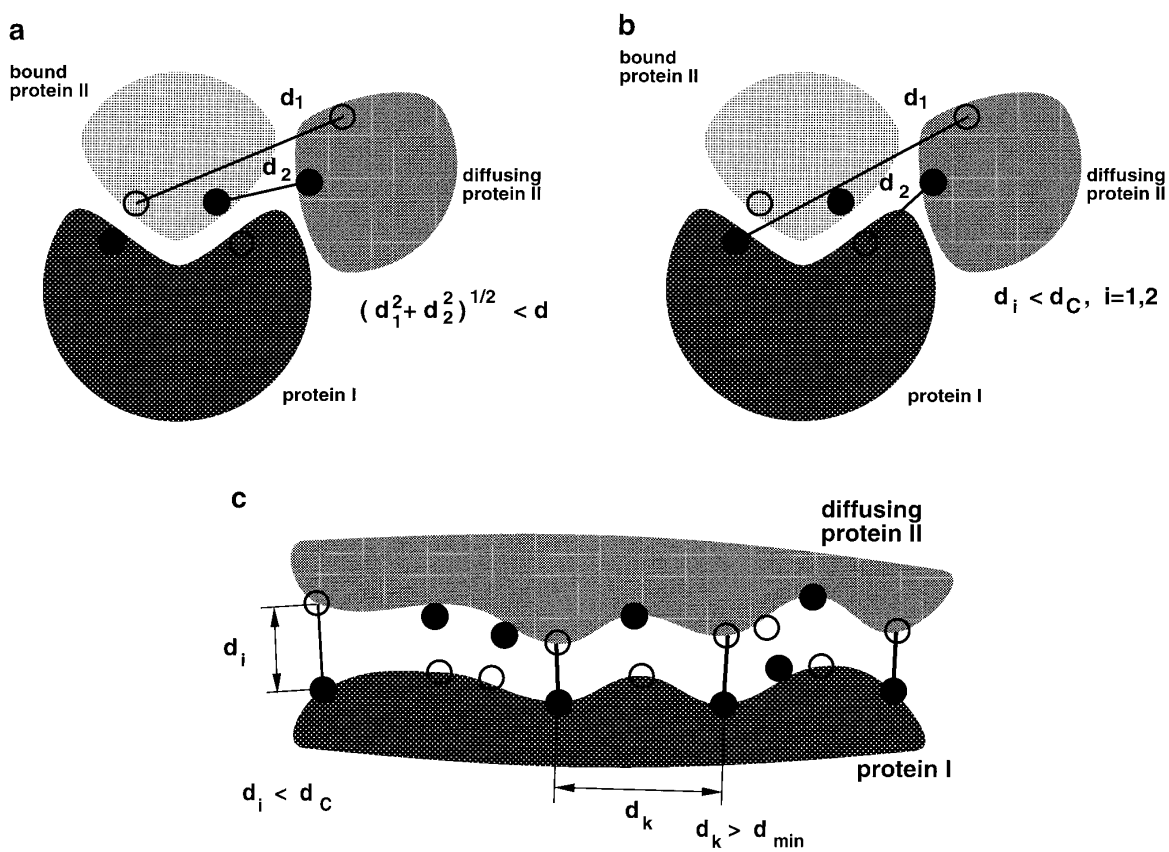

FIG. 3. Schematic representation of different reaction criteria for monitoring encounter complex formation. (a) Criterion based on rms distance of selected atoms of protein II to their position in bound complex. (b) Criterion based on the number of residue contacts. In both (a) and (b), any number of atoms may be used and two in each protein are shown for simplicity. (c) Atomic contact criterion. See text for details.

and run until protein II escapes from the q surface. The number of reaction events during each simulation is counted. Simulations can straightforwardly be run in parallel on multiple processors.

j. Compute Rates and Analyze Trajectories

This is done for diffusional pathways and important regions of the proteins for diffusional encounter.

Comparison of Simulation of Protein–Protein and Protein–Small Molecule Encounters

The theory and the general procedure for computing association rates are the same for protein–protein interactions as for protein–small molecule interactions. However, there are some significant differences that make the association of two macromolecules more difficult to simulate than the association of a macromolecule with a much smaller molecule:

—There are more degrees of freedom. Rotational diffusion must be simulated for protein–protein association, whereas small molecules can often be treated as either a single sphere or two spheres (a dumbbell) covalently bound together that are treated as separate hydrodynamic objects connected by a constrained bond (48).

—The low dielectric of both molecules must be treated; i.e., the test charge model is a worse approximation for protein–protein association than protein–small molecule association.

—Reaction criteria may be harder to define as there is less likely to be a well-defined concave binding site for a protein than for a small molecule.

—HIs between two similar-size macromolecules are likely to have more impact on their association rate than HIs between molecules of very different sizes.

—While internal flexibility in a substrate may be of negligible importance for diffusional encounter with an enzyme, intramolecular motions in proteins are more likely to be important.

Software for BD Simulations

Several programs are available for performing BD simulations, each with different capabilities:

—Macrodox (http://pirn.chem.tntech.edu/macrodox.html), developed by Scott Northrup and colleagues, may be used to perform BD simulations of the bimolecular association with simple and atomic-detail models. It was designed particularly for performing BD simulations of the association of electron transfer heme proteins. Its capabilities also include Tanford–Kirkwood calculations to estimate pK_a values and assign partial charges and numerical solution of the linear and full Poisson–Boltzmann equations.

—UHBD (http://chemcca10.ucsd.edu/~jmbriggs/uhbd.html) (11), developed by Andrew McCammon and colleagues, is designed for BD simulation of enzyme–substrate encounter to compute steady-state and non-steady-state association rates. It can be used for simulations of protein–protein association when one protein is treated with a simplified rather than atomic-detail model. It also has modules for numerical solution of the finite-difference linear and full Poisson–Boltzmann equations, electrostatic binding free energies, stochastic dynamics, molecular mechanics energy minimization, and modeling internal protein flexibility in BD simulations.

—SDA (http://www.embl-heidelberg.de/ExternalInfo/wade/pub/soft/sda.html) was developed by us to be generally applicable to the simulation of diffusional association of biomolecules with simple or atomic-detail models. Additional capabilities include a module (ECM) for computation of effective charges using potential grids computed with the UHBD program and another module to estimate electrostatic enhancement of bimolecular rates by the Boltzmann factor analysis method of Zhou (49).

EXAMPLE APPLICATION: DIFFUSIONAL ENCOUNTER OF BARNASE AND BARSTAR

The association of barnase, an extracellular ribonuclease, with its intracellular inhibitor, barstar, provides a particularly well-characterized example of electrostatically steered diffusional encounter between proteins. The association rate is very fast ($\sim 10^8$–10^9 M^{-1} s^{-1} at 50 mM ionic strength), and studies of the effects of mutation and variation of ionic strength clearly show the influence of electrostatic interactions (7). We have shown that the ionic strength dependence of the rate for the two wild-type proteins and the rates for the wild-type proteins and 11 mutants at 50 mM ionic strength can be reproduced to within a factor of 2 by BD simulation (36). These simulations also provided insights into the structure of the diffusional encounter complex in which barstar tends to be shifted from its position in the crystal structure of the complex toward the guanine-binding loop on barnase.

Simulations of the association of barnase with barstar were carried out using the SDA software following the procedure described above. Many (10,000 per protein variant) trajectories were generated and the conformations sampled in one trajectory are shown in Fig. 4. Here, we draw attention to methodological aspects, correct treatment of which is particularly important for obtaining agreement with experiment.

Electrostatic Model

The importance of accounting for the dielectric heterogeneity and ion exclusion volumes of both proteins

is seen by comparing the results obtained with an effective charge model with those obtained with a test charge model. As shown in Fig. 5, the use of test charges to compute electrostatic forces leads to underestimation of the steering forces. 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, but is modeled as only a 10-fold drop with test charges. The ionic strength dependence of the rate is most severely underestimated at ionic strengths of 300 to 500 mM, implying that the test charge interactions at these and higher ionic strengths are very small at the encounter contact distances. On the other hand, with effective charges, ionic strength dependence is reproduced from 50 to 500 mM ionic strength and interactions are significant at the higher ionic strengths. This shows that the association mechanism is the same at all salt concentrations in this range and that the ionic strength dependence of the rates may be ascribed solely to changes in the electrostatic steering forces.

Definition of Formation of Diffusional Encounter Complex

Reaction criteria based on geometry, energy, and intermolecular contacts were tested for this system as follows:

Geometric Criterion

Two atoms on the binding face of barstar were required to approach within a specified rmsd, d, of their positions in the crystallographically observed complex, as shown in Fig. 3a. This criterion could reproduce the effects on rate for only some mutants. Moreover this was at an rmsd d of 6.5 Å, which is too large a distance to be a realistic measure of the distance at which short-range interactions between the proteins become strong enough to ensure that complexation occurs subsequent to formation of the diffusional encounter complex.

Energetic Criterion

The electrostatic interaction energy was required to be more favorable than a defined threshold. This criterion does not reproduce differences between mutants with a single threshold and is not recommended because it leads to an overestimate of the effects of electrostatic steering which are effectively counted twice.

Residue Contact Criterion

A defined number of correct residue–residue contacts between the two proteins were required in which the distance between specified atoms in the residues was shorter than a defined length d_C, as shown in Fig. 3b. Contact pairs were defined to be as mutually independent as possible. Formation of two contacts shorter

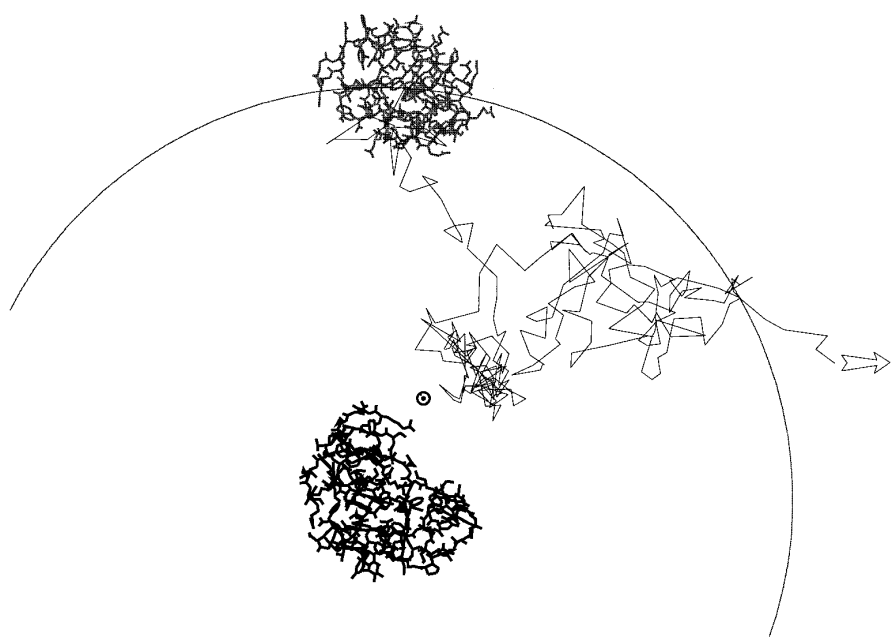

FIG. 4. Trajectory simulated by BD of barstar diffusing toward its binding site on barnase. The trajectory is started with barstar placed at a random position on the sphere shown with a radius of 100 Å centered on barnase. The trajectory is about 68 ns long and the positions of the center of the diffusing barstar are plotted at approximately 200-ps time intervals. The position of the center of barstar in the crystal structure of the complex with barstar is indicated by the dot in the small circle. Most trajectories are shorter than the one shown and in most, barstar does not reach the barnase binding site before diffusing away.

than 5.25 Å or three contacts shorter than 7.5 Å was found to give computed rates in good agreement with experimental data and reproduce the effects of ionic strength (see Fig. 5) (36). The distance of 5.25 Å can be interpreted as a direct, rather than solvent-separated, interaction distance between atoms. A better fit for the diverse set of mutants studied was obtained at three times the experimentally measured association rate with $d_C = 6.25$ Å for two contacts (see Fig. 6), indicating that BD is better suited to model the formation of the transition state than the bound complex.

Atom Contact Criterion

This is similar to the above criterion but the atom–atom contacts are assigned in a fully automated way independent of which residue each atom is in (see Fig. 3c). The contacts are between hydrogen bond donor and acceptor atoms. The specific contacts are selected so that they are approximately equivalent in terms of their contribution to the bimolecular interaction energy. This is achieved by requiring donor–acceptor pairs to be separated by distances greater than d_{min} (6 Å). A contact between two side chains is counted as one contact even if there are two or three contacts between donor and acceptor atoms in the side chains. Potential contacts are defined by, first, tabulating all intermolecular donor–acceptor atom contacts shorter than a distance d_{max} (5 Å) in the experimental structure of the bimolecular complex and, second, generating a table listing dependent pairs of contacts, i.e., those contacts in which either atom is within d_{min} of an atom of the other contact within the same protein. During a BD trajectory, donor–acceptor atom contacts shorter than d_C are monitored. Then the number of independent contacts is computed by ensuring that no more than one contact from any dependent pair of contacts is counted.

With this definition, computed rates for barnase–barstar association agreed well with experimental rates when two contacts shorter than $d_C = 4.5$ Å were required. Computed rates were more accurate than those obtained for the residue contact criterion for variants in which mutation affected the number of residue contacts monitored with the residue contact criterion.

Sensitivity of Model

The absolute magnitude of the rate computed is much more sensitive to modifications in parameters or protein structure than the relative rates for the set of mutants or the ionic strength dependence. The important effects of reaction criteria and charge model on rates computed has been discussed above. Modification of the following additional features of the model altered the absolute rates at 50 mM ionic strength, but did not produce any significant change in the correlation between computed and experimental rates for the set of mutants studied.

• A factor of up to 10 decrease in rates could be obtained by using the coordinates of the unbound barnase from either the crystal structure or the ensemble of structures from the NMR determination, instead of coordinates from the crystal structure of the barnase–barstar complex.

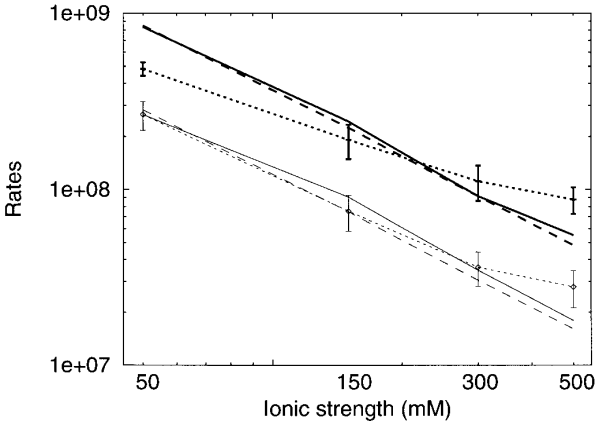

FIG. 5. Ionic strength dependence of the barnase–barstar association rate: ---, Experimentally measured values (plain) and experimentally measured values multiplied by 3 (boldface); · · ·, values computed with test charges; —, values computed with effective charges. For defining encounter complex formation, two contacts were required with contact distances less than 6.5 Å (lower line, plain) and 7.5 Å (upper line, boldface) for test charges and 5.25 Å (lower line, plain) and 6.00 Å (upper line, boldface) for effective charges. Note that the commonly used test charge approximation results in an underestimate of the ionic strength dependence and that this can be corrected by using effective potential-derived charges.

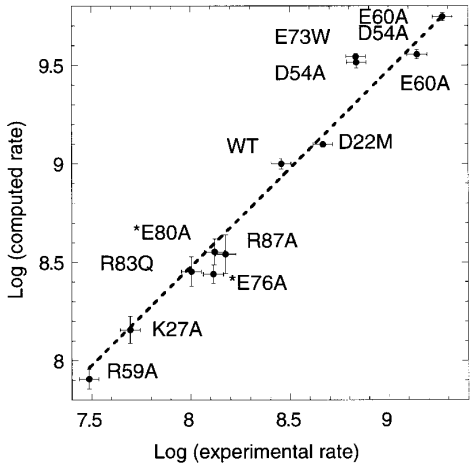

FIG. 6. Comparison of experimental and computed association rates for wild-type barnase and barstar and 11 mutants. A contact distance of 6.25 Å was required when computing the rates. The dashed line presents the dependence log(computed rate) = 3*log(experimental rate).

• A factor of 2 decrease in rates could be obtained by removing the 2-Å-thick Stern layer, or by minimizing hydrogen atom positions with the proteins separated rather than positioned in their bimolecular complex, or by using snapshots from a molecular dynamics simulation of the barnase–barstar complex rather than the crystallographic coordinates.

• No noticeable difference was observed for this system when the nonlinear Poisson–Boltzmann equation was used instead of the linearized equation to compute electrostatic forces.

In summary, the application to barnase–barstar association shows the following:

• The BD simulation method can reproduce experimental rates well. The accuracy is such that BD simulations should be useful for the design of mutants with altered on-rates.

• In the encounter complex (transition state) for this enzyme, two correct contacts are formed. Rotational freedom is not yet fully lost (see Fig. 7).

• Some residues are more important than others for steering binding. Barstar tends to bind first to residues in the guanine-binding loop of barnase and their mutation has a greater impact on association rates than other barnase residues in the binding interface.

CONCLUDING REMARKS

Besides the barnase–barstar case described above, BD simulation has been used to compute protein–protein association rates in a range of applications: electron transfer heme proteins, antigen–antibody binding, and protein self-association (17). The first applications using detailed protein models related to the heme protein cytochrome *c* and its electron transfer partners, cytochrome *c* peroxidase (44) and cytochrome b_5 (45). Electron transfer rates were computed by coupling models describing the electron transfer event to diffusional encounter trajectories of the proteins. Ionic strength dependencies were reasonably reproduced, qualitative insights into the effects of mutations were obtained, and structural information about encounter complexes was derived. For cytochrome *c*–cytochrome *c* peroxidase association, two distinct binding modes were obtained, each containing many alternative conformations, indicating a multitude of electron transfer orientations rather than a single dominant complex. After the simulations of cytochrome *c*–cytochrome *c* peroxidase association were performed, the crystal structure (50) was solved but showed specific complex formation. This apparent discrepancy may arise because the crystal structure should represent the unique bound state, whereas the BD simulations model transi-

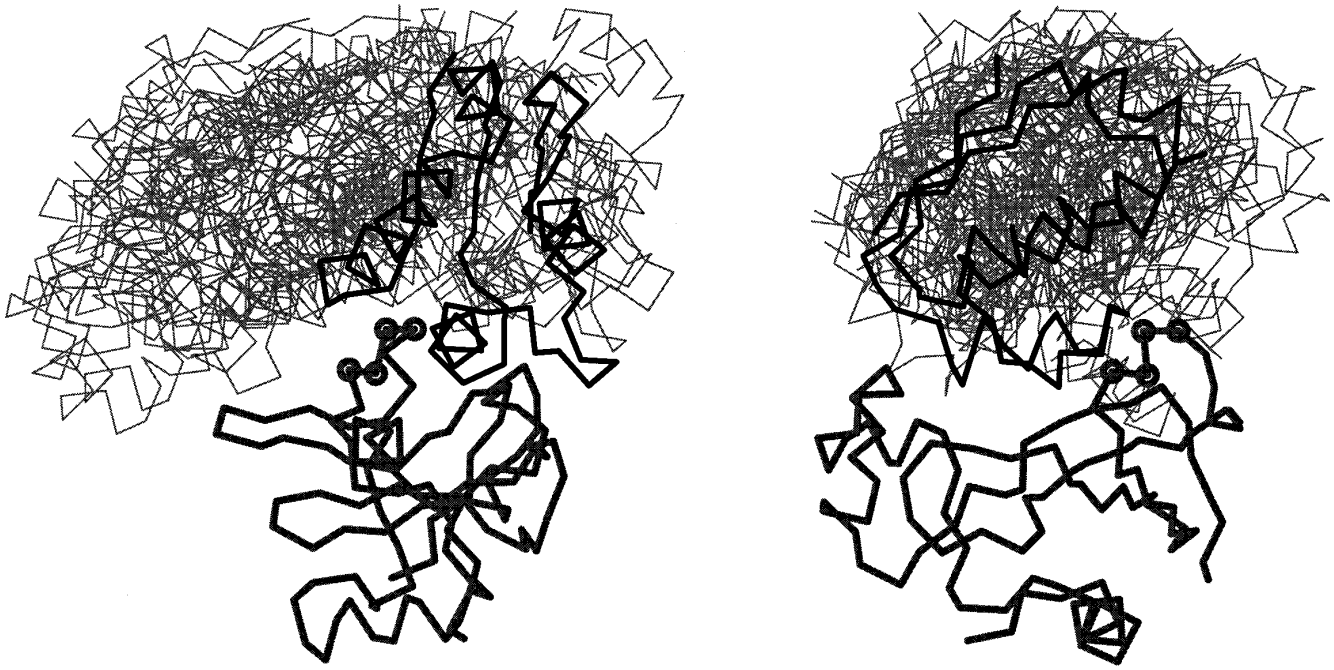

FIG. 7. Encounter complexes for barnase–barstar association. Barnase and barstar are shown in boldface at their positions in the crystal structure of their complex. The gray lines show 17 encounter positions of barstar obtained in different trajectories. The two views differ by a 90° rotation about the vertical axis. Residues 57–60 in the guanine-binding loop of barnase are indicated by circles.

tion state encounter complexes. A similar situation exists for the barnase–barstar system discussed in the previous section.

The importance of electrostatic effects has been observed in BD simulations of antibody–antigen association and protein self-association. Antibody–antigen association rates are typically of the order of 10^6 M^{-1} s^{-1} at physiological ionic strength, yet BD simulations (51, 52) showed that the association rates of lysozyme with the HyHEL-5 antibody are sensitive to ionic strength and mutation of charged residues near the antibody's binding site. The bimolecular association rates of BPTI molecules have been simulated by BD (53) and compared with the kinetics of disproportionation of disulfide groups on the protein surface (54). Association rates are $\sim 10^7$ M^{-1} s^{-1} at 200 mM ionic strength. They increase with ionic strength, as would be expected for molecules of like charge, although some other proteins undergoing self-association display different ionic strength dependencies (54). BD simulations were performed with several models but only with the most realistic model with all atoms, and 128 charges was the computed ionic strength dependence of the rate close to that observed experimentally. A test charge model was used and internal motions were neglected, but these approximations were expected to be partially compensating for this system (53).

In the future, applications can be expected to the association of many other proteins including, for example, those involved in signal transduction, and toxin proteins binding to pore proteins and enzymes. BD simulations of protein association are complementary to other methods to estimate the effects of electrostatic interactions on association rates such as the Boltzmann factor analysis of Zhou (49, 55), which gives estimates of the electrostatic enhancement of on-rates in qualitative agreement with those obtained by BD. It is also complementary to computational methods to dock proteins and predict their bound complexes. Indeed, in addition to providing computed bimolecular rates, BD simulations may provide good starting structures for docking procedures.

The SDA software suite for BD simulations is available from the authors.

ACKNOWLEDGMENT

The authors thank Dr. Indira Shrivastava for reading the manuscript.

REFERENCES

1. Calef, D. F., and Deutch, M. (1983) *Annu. Rev. Phys. Chem.* **34,** 493–524.
2. Berg, O. G., and vonHippel, P. H. (1985) *Annu. Rev. Biophys. Biophys. Chem.* **14,** 131–160.
3. DeLisi, C. (1980) *Q. Rev. Biophys.* **13,** 201–230.
4. Szabo, A., Stolz, L., and Granzow, R. (1995) *Curr. Opin. Struct. Biol.* **5,** 699–705.
5. Malmqvist, M. (1993) *Curr. Opin. Immunol.* **5,** 282–286.
6. Stone, S. R., Dennis, S., and Hofsteenge, J. (1986) *Biochemistry* **28,** 6857–6863.
7. Schreiber, G., and Fersht, A. R. (1996) *Nature Struct. Biol.* **3,** 427–431.
8. Gosting, L. J. (1956) *Adv. Protein Chem.* **11,** 429–554.
9. Johnstone, R. W., Andrew, S. M., Hogarth, M. P., Pietersz, G. A., and McKenzie, I. F. C. (1990) *Mol. Immunol.* **27,** 327–333.
10. Ito, W., Yasui, H., and Kurosawa, Y. (1995) *J. Mol. Biol.* **248,** 729–732.
11. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., and McCammon, J. A. (1995) *Comp. Phys. Commun.* **91,** 57–95.
12. Wade, R. C. (1996) *Biochem. Soc. Trans.* **24,** 254–259.
13. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) *Nature* **358,** 347–351.
14. Northrup, S. H., and Erickson, H. P. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3338–3342.
15. Einstein, A. (1905) *Ann. Phys. (Leipzig)* **17,** 549.
16. Smoluchowski, M. V. (1906) *Ann. Phys. (Leipzig)* **21,** 756.
17. Madura, J. D., Briggs, J. M., Wade, R. C., and Gabdoulline, R. R. (1997) Encyclopedia of Computational Chemistry, in press.
18. Ermak, D. L., and McCammon, J. A. (1978) *J. Chem. Phys.* **69,** 1352–1360.
19. Smoluchowski, M. V. (1917) *Z. Phys. Chem.* **92,** 129–168.
20. Northrup, S. H., Allison, S. A., and McCammon, J. A. (1984) *J. Chem. Phys.* **80,** 1517–1524.
21. Huber, G. A., and Kim, S. (1996) *Biophys. J.* **70,** 97–110.
22. Zhou, H.-X. (1990) *J. Phys. Chem.* **94,** 8794–8800.
23. Luty, B. A., McCammon, J. A., and Zhou, H.-X. (1992) *J. Chem. Phys.* **97,** 5682–5686.
24. Zhou, H.-X. (1993) *Biophys. J.* **64,** 111–1726.
25. Luty, B. A., Wade, R. C., Madura, J. D., Davis, M. E., Briggs, J. M., and McCammon, J. A. (1993) *J. Phys. Chem.* **97,** 233–237.
26. Northrup, S. H., Reynolds, J. C. L., Miller, C. M., Forrest, K. J., and Boles, J. O. (1986) *J. Am. Chem. Soc.* **108,** 8162–8170.
27. Davis, M. E., and McCammon, J. A. (1989) *J. Comp. Chem.* **10,** 386–391.
28. Gabdoulline, R. R., and Wade, R. C. (1996) *J. Phys. Chem.* **100,** 3868–3878.
29. DelaTorre, J. G., and Bloomfield, V. A. (1981) *Q. Rev. Biophys.* **14,** 81–139.
30. Antosiewicz, J., Gilson, M. K., Lee, I. H., and McCammon, J. A. (1995) *Biophys. J.* **68,** 62–68.
31. Wolynes, P. G., and Deutch, J. M. (1976) *J. Chem. Phys.* **65,** 450–454.
32. Friedman, H. L. (1966) *J. Phys. Chem.* **70,** 3931–3933.
33. Allison, S. A., Srinivasan, N., McCammon, J. A., and Nortrup, S. H. (1984) *J. Phys. Chem.* **88,** 6152–6157.
34. Deutch, J. L., and Felderhof, B. U. (1973) *J. Chem. Phys.* **59,** 1669–1671.
35. Brune, D., and Kim, S. (1994) *Proc. Natl. Acad. USA* **91,** 2930–2934.

36. Gabdoulline, R. R., and Wade, R. C. (1997) *Biophys. J.* **72,** 1917–1929.

37. Wade, R. C., Davis, M. E., Luty, B. A., Madura, J. D., and McCammon, J. A. (1993) *Biophys. J.* **64,** 9–15.

38. Demchuk, E., and Wade, R. C. (1996) *J. Phys. Chem.* **100,** 17373–17387.

39. Hooft, R. W., Sander, C., and Vriend, G. (1996) *Proteins* **26,** 363–376.

40. Jorgensen, W., and Tirado-Rives, J. (1988) *J. Am. Chem. Soc.* **110,** 1657–1666.

41. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217.

42. Sharp, K. A., and Honig, B. (1990) *Annu. Rev. Biophys. Chem.* **19,** 301–332.

43. Northrup, S. H., Boles, J. O., and Reynolds, J. C. L. (1987) *J. Phys. Chem.* **91,** 5991–5998.

44. Northrup, S. H., Boles, J. O., and Reynolds, J. C. L. (1988) *Science* **241,** 67–70.

45. Northrup, S. H., Thomasson, K. A., Miller, C. M., Barker, P. D., Eltis, L. D., Guillemette, J. G., Inglis, S. C., and Mauk, A. G. (1993) *Biochemistry* **32,** 6613–6623.

46. Antosiewicz, J., and Porschke, D. (1989) *J. Phys. Chem.* **93,** 5301.

47. Brune, D., and Kim, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 3835–3839.

48. Antosiewicz, J., Briggs, J. M., and McCammon, J. A. (1996) *Eur. Biophys. J.* **24,** 137–141.

49. Zhou, H.-X. (1996) *J. Chem. Phys.* **105,** 7235–7237.

50. Pelletier, H., and Kraut, J. (1992) *Science* **258,** 1748–1755.

51. Kozack, R. E., and Subramaniam, S. (1993) *Protein Sci.* **2,** 915–926.

52. Kozack, R. E., d'Mello, M. J., and Subramaniam, S. (1995) *Biophys. J.* **68,** 807–814.

53. Nambi, P., Wierzbicki, A., and Allison, S. A. (1991) *J. Phys. Chem.* **95,** 9595–9600.

54. Sommer, J., Jonah, C., Fukuda, R., and Bersohn, R. (1982) *J. Mol. Biol.* **159,** 721.

55. Zhou, H.-X., Briggs, J. M., and McCammon, J. A. (1996) *J. Am. Chem. Soc.* **118,** 13069–13070.