JCTC Journal of Chemical Theory and Computation

Free Energy Landscape of Protein−Protein Encounter Resulting from Brownian Dynamics Simulations of Barnase:Barstar

Alexander Spaar and Volkhard Helms*

Center for Bioinformatics, Saarland University, Im Stadtwald, D-66041 Saarbrücken, Germany

Received February 24, 2005

Abstract: Over the past years Brownian dynamics (BD) simulations have been proven to be a suitable tool for the analysis of protein−protein association. The computed rates and relative trends for protein mutants and different ionic strength are generally in good agreement with experimental results, e.g. see ref 1. By design, BD simulations correspond to an intensive sampling over energetically favorable states, rather than to a systematic sampling over all possible states which is feasible only at rather low resolution. On the example of barnase and barstar, a well characterized model system of electrostatically steered diffusional encounter, we report here the computation of the 6-dimensional free energy landscape for the encounter process of two proteins by a novel, careful analysis of the trajectories from BD simulations. The aim of these studies was the clarification of the encounter state. Along the trajectories, the individual positions and orientations of one protein (relative to the other) are recorded and stored in so-called *occupancy maps*. Since the number of simulated trajectories is sufficiently high, these occupancy maps can be interpreted as a probability distribution which allows the calculation of the entropy landscape by the use of a locally defined entropy function. Additionally, the configuration dependent electrostatic and desolvation energies are recorded in separate maps. The free energy landscape of protein−protein encounter is finally obtained by summing the energy and entropy contributions. In the free energy profile along the reaction path, which is defined as the path along the minima in the free energy landscape, a minimum shows up suggesting this to be used as the definition of the *encounter state*. This minimum describes a state of reduced diffusion velocity where the electrostatic attraction is compensated by the repulsion due to the unfavorable desolvation of the charged residues and the entropy loss due to the increasing restriction of the motional freedom. In the simulations the orientational degrees of freedom at the encounter state are found to be less restricted than the translational degrees of freedom. Therefore, the orientational alignment of the two binding partners seems to take place beyond this free energy minimum. The free energy profiles along the reaction pathway are compared for different ionic strength and temperature. This novel analysis technique facilitates mechanistic interpretation of protein−protein encounter pathways which should be useful for interpretation of experimental results as well.

Introduction

The association of two proteins is commonly described by a simple picture where initially the free proteins diffuse randomly in solution until they enter into the electrostatic field of the other, the so-called steering region. In this region they move by directional diffusion toward the so-called *encounter complex*. Once the encounter complex is formed, it can either dissociate or evolve into the final complex. The process of association to finally form a bound complex can

* Corresponding author phone: +49 681 302 64165; e-mail: volkhard.helms@bioinformatik.uni-saarland.de.

10.1021/ct050036n CCC: $30.25
Published on Web 05/18/2005

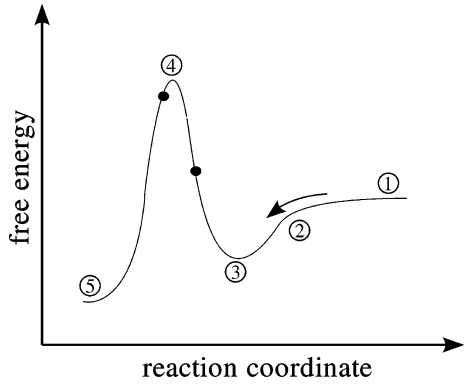

Figure 1. Sketch of the free energy profile along the reaction pathway: (1) denotes the region of free diffusion, (2) the steering region, (3) a minimum of free energy, here used as a definition of the encounter complex, (4) the transition state, and (5) the bound state. The region up to the first black circle, including the steering region and the formation of the encounter complex, can be well modeled by BD simulations. The second circle indicates the point-of-no-return, i.e., where the proteins are committed to form a bound complex.

thus be divided into two steps, where the first one is the formation of the encounter complex, the end-point of the diffusional encounter, while in the second step the proteins subsequently rearrange to form the final bound complex.[2] Figure 1 shows a sketch of the free energy profile along the path of minimal free energy, also called reaction pathway, for an electrostatically steered association: (1) denotes the region of free diffusion, (2) the steering region, (3) a minimum of free energy, here used as definition of the encounter complex, (4) the transition state, and (5) the bound state. Since the first process including the formation of the encounter complex is diffusion limited, it can be modeled by Brownian Dynamics (BD) simulations. The computed association rates from these simulations are in excellent agreement with experimental results once a definition of the so-called *diffusional encounter complex* (which is different from our definition) is used.[1] This diffusional encounter complex is supposed to be located just behind the transition state (4). In the transition state there is a 50:50 chance of binding or association; a position closer than the transition state reflects a 100% probability of binding.

Typical BD simulations of protein−protein encounter employ a series of simplifications: the proteins are modeled as rigid bodies and short-range interactions such as van der Waals forces and the formation of hydrogen bonds and salt bridges are not modeled. However, both aspects become important when the proteins approach the bound conformation, i.e., in the region beyond the diffusion limited regime. To model this second step in the process of association, the transition from the encounter to the bound state, likely more detailed simulation methods have to be applied. In Figure 1 the region up to the first black circle, including the steering region and the formation of the encounter complex, can be well modeled by BD simulations. The second circle indicates the point-of-no-return, i.e., where the proteins are committed to form a bound complex.

While the above-described two-step model of protein−protein association certainly holds true, the detailed steps in diffusional protein−protein encounter are still subject of discussion. In particular the definition of the encounter complex, the intermediate step before the formation of the bound complex, is not completely clear. In a recent article[3] Gabdoulline and Wade have summarized several structural approaches, e.g. the formation of two or three contacts with a certain maximum distance (set by comparison with experimental results),[4] the rotational restriction and the surface area buried in the protein−protein interface,[5] or a maximum deviation of two orientational angles from the bound structure.[6] On the other hand, Camacho and co-workers[7] suggested defining the encounter state as a local minimum in the free energy landscape, an approach which has been taken up in this work. Since protein−protein encounter can be understood as diffusional motion on a funnel-shaped potential energy surface, the knowledge of the free energy landscape should provide a full picture about the encounter process. In their work, the free energy landscape, which was defined as the sum of the electrostatic and desolvation energy, was determined by a systematic sampling of the 6-dimensional conformation space. A disadvantage of this systematic sampling is the immense computational effort if it is performed even at medium resolution. For example, a resolution of 10° for the five angular coordinates and 5 Å for the radial coordinate in a range of 50 Å results in a computation of $36^3 * 18^2 * 10 \sim 10^8$ values. An attractive alternative are simulation methods that give a preferential sampling of the energy landscape, i.e., along the energy valleys.

Here, we present results from BD simulations for the example of barnase and barstar with the aim of clarifying the definition of the encounter state. The ribonuclease barnase and its inhibitor barstar provide a well characterized model system of electrically steered diffusional encounter.[1,5−10] In our studies we examined the electrostatic and desolvation energy along the protein trajectories and additionally recorded the occupancy of the configuration space. This additional information provides insight into the restrictions of the positional and orientational degrees of freedom during the encounter process. These occupancy maps are interpreted as probability distributions in order to derive the entropy landscape by applying a locally defined entropy function. This entropy function takes into account the occupancies of all those configurations, which are reachable from the particular position and orientation within one Brownian dynamics time step.

Estimations of the loss of translational and rotational entropy during the complex formation of barnase and barstar using geometric constraints gave values around 15 kcal/mol,[5] while the application of the gas-phase expressions of the translational and rotational entropy results in 28 kcal/mol.[11] In contrast, Frisch and co-workers showed in an experimental study that the activation entropy at the transition state is close to zero.[12] In this case the entropy corresponds to the loss of translational, rotational, and internal degrees of freedom as well as the entropy of the surrounding solvent, i.e., neither of these processes is very pronounced, or their contributions cancel each other. These examples show that the issue of

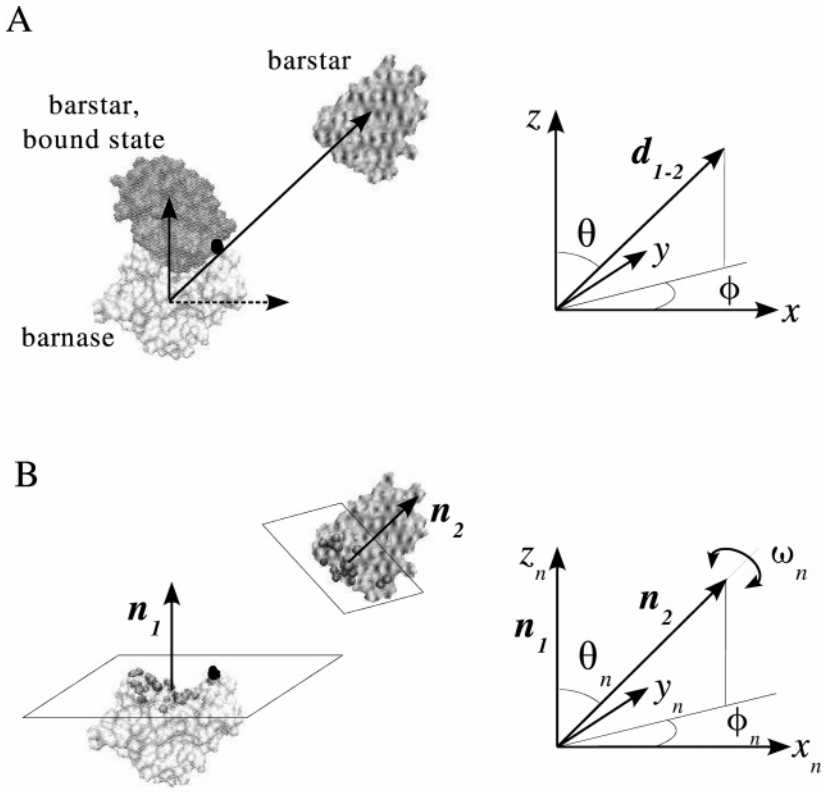

Figure 2. Definition of the reference coordinate system: (A) the positional coordinates and (B) the orientational coordinates. Barnase (protein 1) is displayed in light gray, barstar (protein 2) in dark gray. The black points indicate the position of SER38:OG used for the definition of the *x*-axis. The definition of the coordinates and the angles is given in the text. Figures 2 and 8 were generated using the VMD program.[28]

deriving the entropy for protein−protein association is controversial.

As a final point of our analysis, the free energy landscape is computed by summing the electrostatic and desolvation energies as well as the translational and rotational entropy losses. The free energy profile along the reaction path, which is defined as the path along the minima of the free energy landscape, contains detailed information about the encounter process. The characteristic minimum in the free energy profile is located at the intersection of a favorably decreasing electrostatic interaction energy with an unfavorable desolvation energy and, most notably, with the unfavorable restriction of translational and rotational degrees of freedom. This minimum should be amenable to time-resolved spectroscopic investigation, possibly only in the future.

Methods

Structures. We used the coordinates of barnase and barstar (see Figure 2) as they were prepared by Gabdoulline and Wade and included in the SDA software package.[10]

In their work, the coordinates of the barnase-barstar complex, determined to 2.0 Å resolution by Buckle and coworkers,[8] were taken from the protein data bank (PDB code 1brs). For barnase chain A was used and for barstar chain D. The two other pairs of chains in the crystal structure were used to model the missing side chain atoms in the A and D chains. Crystallographic water molecules were removed. Polar hydrogen atoms were added, and their positions were optimized by energy minimization with the CHARMM program[13] using the QUANTA molecular graphics package.[14] Side-chain conformations were kept the same as in the barnase−barstar complex.

Computation of Forces. In the BD simulations of this work the computation of forces and torques which act on the proteins are performed as it is described in ref 15. First we give a brief overview of the steps involved: For each protein electrostatic potential grids are computed, taking into account the inhomogeneous dielectric medium and the surrounding ionic solvent. A set of effective charges is computed for each protein which represent the external electrostatic potential of the molecule. These sets of effective charges are used to calculate the intermolecular electrostatic interactions during the BD simulation. A further contribution is computed for the interaction of the charges of the proteins with precomputed desolvation grids representing the penalty due to charge desolvation by the low dielectric of the protein interior. Short-range forces are treated in a rather coarse manner by prohibiting overlap of the exclusion volumes of the proteins. The results of our work, the analysis of the encounter pathways and the encounter free energy profiles, are complementary to recent studies of association rates of barnase and barstar, in particular to the work of R. R. Gabdoulline and R. C. Wade. To allow a comparison to their results, we used the same parameters for the computation of forces as in ref 1.

Now the individual steps are described in detail. For the modeling of the long-range electrostatic interaction of the proteins, solutions of the full Poisson−Boltzmann equation

Table 1: Shown Are the Solvent Dielectric ϵ_{solv}, Viscosity η, and Diffusion Coefficients for Barnase and Barstar for the Temperatures Used in the Simulations[a]

T [K]	ϵ_{solv}	η [Pa s]	D [Å^2/ps]	$D_{R,bn}$ [rad^2/ps]	$D_{R,bs}$ [rad^2/ps]
200	122.4	4.01×10^{-2}	4.40×10^{-4}	5.87×10^{-7}	6.60×10^{-7}
220	111.9	1.39×10^{-2}	1.39×10^{-3}	1.86×10^{-6}	2.09×10^{-6}
240	102.2	5.78×10^{-3}	3.67×10^{-3}	4.89×10^{-6}	5.50×10^{-6}
260	93.3	2.74×10^{-3}	8.37×10^{-3}	1.12×10^{-5}	1.26×10^{-5}
280	85.2	1.45×10^{-3}	1.71×10^{-2}	2.28×10^{-5}	2.56×10^{-5}
300	77.7	8.33×10^{-4}	3.18×10^{-2}	4.24×10^{-5}	4.77×10^{-5}
320	70.9	5.13×10^{-4}	5.51×10^{-2}	7.34×10^{-5}	8.26×10^{-5}
340	64.7	3.35×10^{-4}	8.97×10^{-2}	1.20×10^{-4}	1.35×10^{-4}
360	59.0	2.29×10^{-4}	1.39×10^{-1}	1.85×10^{-4}	2.08×10^{-4}
380	53.8	1.63×10^{-4}	2.06×10^{-1}	2.75×10^{-4}	3.09×10^{-4}
400	49.0	1.20×10^{-4}	2.94×10^{-1}	3.93×10^{-4}	4.42×10^{-4}

[a] D is the relative translational diffusion coefficient, and $D_{R,bn}$ and $D_{R,bs}$ are the rotational diffusion coefficients for barnase and barstar, respectively.

were computed for each protein using the UHBD program.[16] Partial atomic charges and atomic radii were assigned from the OPLS parameter set.[17] The protonation states of titratable residues were assigned according to their standard protonation states at the experimental pH of 8.0.[9] Grids with dimensions of 150 × 150 × 150 nodes and a 1.0-Å spacing centered on each of the proteins were used. The ionic strength of the solvent was varied in 10 logarithmic steps from 0 to 1600 mM at a constant temperature of 300 K. To model infinite ionic strength one simulation was performed where all the interaction forces except for the exclusion forces were switched off. For the study of the temperature dependence, the ionic strength was set to 50 mM, while the temperature was varied in steps of 20 K in the range between 200 and 400 K. Phase transitions of water at $T = 273$ and 373 K were not considered (see Discussion). In all calculations the dielectric constant for the protein interior was set to 4.0. The external solvent dielectric was calculated by a polynomial fit of the dielectric constant of water for the temperatures as given in ref 18, at 300 K being 77.7. [Besides the polynomial approach $\epsilon(T) = b_0 + b_1T_c + b_2T_c^2 + b_3T_c^3$, $T_c = T - 273.15$ K, an exponential fitting equation $\epsilon(T) = exp(b_0 + b_1T_c)$ is suggested in ref 18. The relative deviation of the fitted to the given values, however, was much smaller for the polynomial approach. Again, phase transitions were omitted.] The values for the solvent dielectric which were used in the simulations are listed in Table 1.

The effective charge method (ECM)[19] was used to derive charges that represent the external electrostatic potential of the molecule in a uniform dielectric medium. The effective charges were fitted to reproduce the electrostatic potential in a 3 Å thick layer starting at the accessible surface defined by a probe of radius 4 Å and extending outward from the protein. To compute forces and torques acting on protein 2(1), the array of effective charges for protein 2(1) is placed on the electrostatic potential grid of protein 1(2).

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 with 0.5 Å grid spacing. If a move during the BD simulation would result in van der Waals overlap, the BD step is repeated with different random numbers until no overlap occurs. The surface-exposed atoms of the smaller protein are listed, and steric overlap is defined to occur when one of the surface-exposed atoms is projected on a grid point which represents the interior of the larger protein.[20]

Charge desolvation penalties are computed in an approximate fashion which treats the solvation of each charge independently.[15] The charge desolvation penalty of one protein is taken as the sum of desolvation penalties of each charge of that protein. The desolvation penalty of each charge is the sum of desolvation penalties due to the low dielectric cavity of each atom of the other protein. The desolvation energy of protein 1, due to the presence of protein 2, is approximated as

$$\Delta G_{ds} = \alpha \frac{\epsilon_s - \epsilon_p}{\epsilon_s(2\epsilon_s + \epsilon_p)} \sum_{ij} (1 + \kappa r_{ij})^2 \, e^{-2\kappa r_{ij}} \frac{q_i^2 a_j^3}{r_{ij}^4} \quad (1)$$

where κ is the Debye−Hückel parameter, ϵ_s and ϵ_p are the dielectric constants of the solvent and the protein, respectively, q_i is the effective charge on the ith atom of protein 1, a_j is the radius of the jth atom of protein 2, and r_{ij} is the distance between the two atoms. The summation is carried out over all possible pairs of effective charges on protein 1 and atoms on protein 2. The scaling factor α for the weighting between electrostatic interaction and desolvation terms was set to 1.67. The validity of this choice may be assessed from the electrostatic interaction free energy calculations shown in Figures 1 and 2 of ref 1.

The atom−atom contacts of the *reaction patches* are assigned in a fully automated way independent of which residue each atom is in ref 10. Possible contacts are those pairs between hydrogen-bond donor and acceptor atoms having a separation distance of less than 5.0 Å in the X-ray structure of the complex.

Brownian Dynamics Simulation. For the BD simulations we used the software package SDA (Simulation of Diffusional Association of proteins)[4,10] which was modified to allow for a detailed analysis of the trajectories. As in other BD simulations of protein−protein encounter like UHBD[16] or MacroDox,[21] SDA makes simplifications which become important only at small protein−protein separations: the proteins are modeled as rigid bodies and short-range interactions as van der Waals forces and the formations of hydrogen bonds and salt bridges are not modeled.

For each set of parameters, 10 000 trajectories were simulated. The trajectories start with the two proteins at a center-to-center distance b with randomly chosen orientations and finish when the proteins reach a center-to-center distance $c > b$. In all simulations b was chosen as 100 Å and c as 500 Å. This corresponds to 7.4 and 36.8 Debye lengths at 300 K and 50 mM. The diffusion equation is solved by the Ermak−McCammon algorithm.[22] The translational Brownian motion of two interacting proteins is simulated as the displacement $\Delta\mathbf{r}$ of the relative separation vector $\mathbf{r}$ during a time step Δt according to the relation

$$\Delta\mathbf{r} = \frac{D\Delta t}{k_B T}\mathbf{F} + \mathbf{R}, \text{ with } \langle\mathbf{R}\rangle = 0 \text{ and } \langle\mathbf{R}^2\rangle = 6D\Delta t \quad (2)$$

where **F** is the systematic interparticle force, k_B is the Boltzmann constant, T is the temperature, and **R** is the stochastic displacement arising from collisions of the proteins with solvent molecules. Two analogous formulas are used to generate the rotational motions of the two proteins in terms of rotation angle $\mathbf{w}_j = (w_{1j}, w_{2j}, w_{3j})$, torque $\mathbf{T}_{ij}$ acting on protein i due to protein j, and rotational diffusion constant D_{iR} of each protein i ($i, j = 1, 2, i \neq j$):

$$\Delta\mathbf{w}_i = \frac{D_{iR}\Delta t}{k_B T}\mathbf{T}_{ij} + \mathbf{W}_i \text{ with } \langle\mathbf{W}_i\rangle = 0 \text{ and } \langle\mathbf{W}_i^2\rangle = 6D_{iR}\Delta t \quad (3)$$

The diffusional properties of the molecules are assumed to be isotropic. In ref 1 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 in aqueous solution at 298 K, and rotational diffusion constants of 4.0×10^{-5} and 4.5×10^{-5} radian^2/ps for barnase and barstar, respectively. Based on these values, we calculated the diffusion coefficients for temperatures in the range from 200 to 400 K according to the Einstein−Stokes relationship $D \propto T/\eta(T)$. Here the temperature dependence of the solvent viscosity is approximately given by the Arrhenius behavior $\eta(T) = Ae^{E_a/RT}$, and E_a is the activation energy for the solvent molecules in their respective cages (water: $A = 1.05 \times 10^{-6}$ Pa s, $E_a = 4.02$ kcal/mol[23]). Table 1 lists the viscosity and the diffusion coefficients for the temperatures used in the simulations.

For $T = 300$ K, the time step was set to 1.0 ps for center-to-center distances up to 75 Å. For larger distances it increased linearly with the intermolecular separation up to about 200 ps at 500 Å. This corresponds to an average random displacement of 0.4 Å at small and medium separations and 6.0 Å at the largest separation of 500 Å. For the simulations at higher or lower temperature the time step was adjusted to keep the average random displacement constant, according to $\Delta r = \sqrt{6D\Delta t}$.

To reduce the computational cost of the simulations, no hydrodynamic interactions were considered. It was shown that the effect on the protein−protein association rates is rather small.[24] Also in previous BD studies with simplified cytochrome *c* molecules it was found that hydrodynamics has only a small influence.[25] Hydrophobic forces are not included, since they are assumed to become important only at small contact distances.

The simulation software SDA also allows for the computation of association rates. This was covered in detail in previous work by others.[1,10,26] In this study, we concentrate on the analysis of the trajectories, largely omitting the computed association rates.

For the simulation at 300 K and 50 mM the computation time for 10 000 trajectories including the trajectory analysis and entropy calculation amounts to about 160 h on one pentium IV processor (2.8 GHz). With an average length of a single trajectory of about 70 000 steps, the average simulated time of one trajectory is 1.5 μs. For the simulations with different ionic strength and temperature these values are similar.

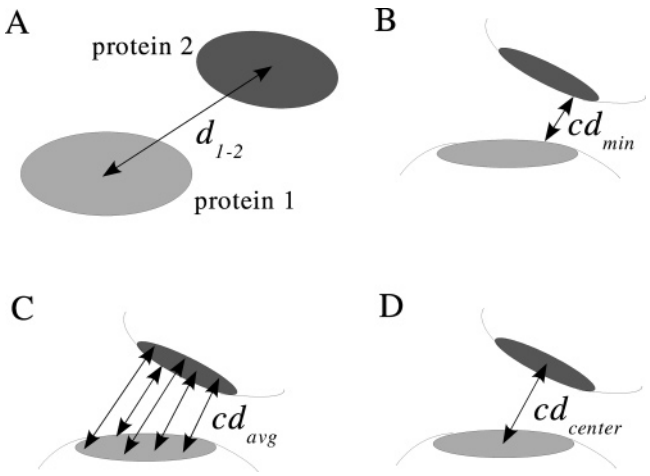

Figure 3. Four possible definitions of the distance axis used here: the center-to-center distance (A), the minimum contact distance (B), the average contact distance (C), and the distance between the geometric centers of the contact surfaces (D).

Trajectory Analysis. After each time step of the simulated trajectories the positional and orientational coordinates of protein 2 (relative to protein 1) are computed with respect to a reference coordinate system of the 6-dimensional configuration space. During the simulations these coordinates are assigned to the nodes of a 6-dimensional grid on which the occupancy, energy, and entropy maps are computed. Note that the maps for the positional and orientational coordinates were computed separately since the handling of a 6-dimensional grid even at moderate resolution is computationally not feasible.

For the positional coordinates a spherical coordinate frame was chosen. If the center-to-center distance d_{1-2} is defined as the reaction coordinate, the origin of the coordinate frame is the center of protein 1. In the case of using a contact distance as reaction coordinate (see below), the origin is the geometric center of the interaction patch of protein 1. The *z*-axis is defined as the vector from the center of protein 1 to the center of protein 2 in the bound state of the crystal structure, see Figure 2A. Without loss of generality, the *x*-axis is defined (orthogonal to the *z*-axis) by the vector from the center of the reaction patch of protein 1 to the first atom in the list of reaction atoms. In the case of barnase this was SER38:OG (shown as black sphere in Figure 2). Therefore, the negative *x*-axis points approximately into the direction from the patch center towards the guanine binding loop. The *y*-axis is defined orthogonal to the *x*- and *z*-axes. The angle between the *z*-axis and the center-to-center vector for a given trajectory position is denoted as the polar angle θ. The azimuthal angle ϕ is the angle in the *xy*-plane from the *x*-axis. For the visualization of the computed occupancy maps (Figure 4A, see below), the coordinates θ and ϕ are replaced by 'semi-Cartesian' coordinates ϕ_x and ϕ_y, which are the *x*- and *y*-contributions of the normalized center-to-center vector.

The orientational coordinates are defined similarly to the positional coordinates (see Figure 2B) by exchanging the center-to-center vectors with the normal vectors of the reaction patches of the proteins (denoted as $\mathbf{n_1}$ and $\mathbf{n_2}$). Here, the normal vector of protein 1 is defined to point outward

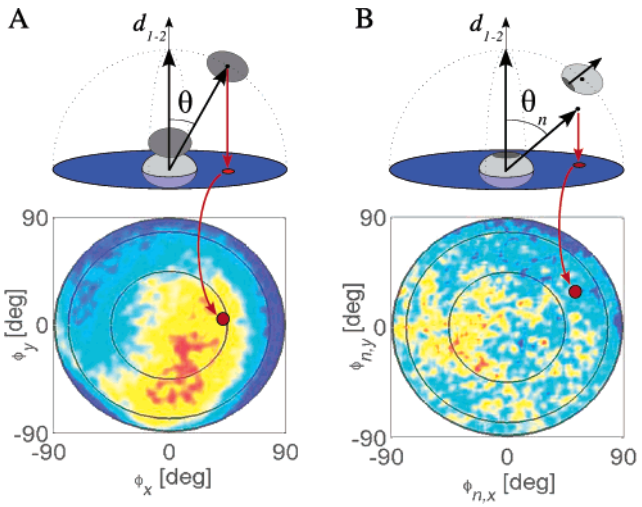

Figure 4. The computation of the occupancy maps, displayed by the projection of the position (A) and orientation (B) of protein 2 (relative to protein 1) onto a plane, which is perpendicular to the center-to-center vector of the proteins in the bound state, and to the normal vector of the reaction patch of protein 1, respectively (see text). The colors in the occupancy maps change from blue (low) to red (high).

of the protein, while the normal vector of protein 2 is defined to point inward, so that in the bound state the two vectors are approximately parallel. The z_n-axis is given by the normal vector of protein 1, and the x_n-axis is defined analogously using the first reaction atom of protein 1. Index n refers to the normal vectors of the reaction patches. The y_n-axis is defined orthogonal to the x_n- and z_n-axis. Again, θ_n is the angle between the vectors and ϕ_n the azimuthal angle. The third orientational coordinate, the angle ω_n, is given by the rotation of protein 2 around the normal vector of the reaction patch, $\omega_n = 0$ denoting the orientation of protein 2 in the bound state. Analogous to the positional coordinates, the orientational coordinates θ_n and ϕ_n are replaced by $\phi_{n,x}$ and $\phi_{n,y}$ for the visualization of the computed occupancy maps (Figure 4B, see below).

When aiming at a global description of the encounter of the two proteins, the center-to-center distance d_{1-2} (Figure 3A) is certainly the ideal definition for the *reaction coordinate* (the third positional coordinate). This definition, however, is independent of the orientation of the proteins with respect to their reaction patches and is therefore problematic for analyzing the encounter of the interaction patches. At closer distances it is therefore preferable to define a so-called 'contact distance'. Gabdoulline and Wade[10] used the definition of the minimum contact distance cd_{min} (Figure 3B), i.e., the minimal distance between all contact pairs. Alternatively, the average value of all contact pair distances, cd_{avg} (Figure 3C), gives a balanced description of the orientation of both interfaces. A further possible definition is the distance between the geometric centers of the protein interfaces, cd_{center} (Figure 3D).

As already mentioned, the occupancy, energy, and entropy maps are computed on a grid while treating the positional and orientational coordinates separately. One set of 3-dimensional matrices is computed for the translational coordinates, a second set for the rotational coordinates $\phi_{n,x}$ and $\phi_{n,y}$ plus the distance coordinate, and a third matrix for ω_n and the distance coordinate.

Computing the occupancy maps with respect to the positional coordinates can be understood as projecting the position of protein 2 onto a plane which is perpendicular to the center-to-center vector d_{1-2} of the proteins in the bound state, see Figure 4A. The origin of these occupancy maps is given by the position of protein 2 in the bound state. After each time step, the occupancy value of the grid element associated with the position of protein 2 is increased by 1. Note that for positions of protein 2 above the denoted plane (for $\theta \leq \pi$) and below (for $\theta > \pi$) two distinct sets of matrices are needed which represent the upper and lower half spheres.

The occupancy maps with respect to the orientational coordinates $\phi_{n,x}$ and $\phi_{n,y}$ are computed similarly. The normal vector of protein 2 is projected onto a plane which is perpendicular to the normal vector of protein 1, see Figure 4B. Here, the origin is the tip of the normal vector of the reaction patch of protein 1. Again, two distinct sets of matrices are needed which represent the upper and lower half spheres.

Figure 5 displays the resulting occupancy maps of the simulation at 50 mM and 300 K. A detailed description is given in the Results section. The electrostatic and desolvation energies as well as the translational and rotational entropy losses (see below) are recorded in further sets of matrices (not shown). In the matrices assigned to the electrostatic and desolvation energies the minimum values for the given position/orientation are stored, thus allowing to finally identify the minimum free energy paths.

The spacing of the grid representing the conformation space should be chosen large enough to ensure a good statistics, but at the same time fine enough to reveal the details of the encounter process, especially in combination with the choice of Δr and Δw in the entropy definition, see below. Along the angles ϕ_x and ϕ_y as well as $\phi_{n,x}$ and $\phi_{n,y}$ we used a discretization with 2×101 nodes (for the upper and lower half spheres), i.e., an average step size of 1.8°. For ω_n we used a step size of 4°, i.e., 90 nodes. Along the distance axis a spacing of 2 Å was used over a distance of 80 Å. The initial value was set by the corresponding distance of the proteins in the bound state. Finally, the occupancy maps were normalized with respect to the size of the volume elements of the grids.

Calculation of the Entropy Landscape. By interpreting the computed occupancy maps as probability distributions, the contribution of the translational and rotational entropy to the free energy landscape is computed by the restriction of the degrees of motional freedom. The total entropy loss of protein−protein encounter is calculated as the sum of the translational and rotational entropy[27]

$$\Delta S = \Delta S_{trans} + \Delta S_{rot} \tag{4}$$

The splitting of the entropy into the translational and rotational parts is valid if the positions and the orientations of the proteins are decoupled. This certainly holds true if the proteins are far apart. Therefore, eq 4 can be assumed to be approximately valid in the region of free diffusion until

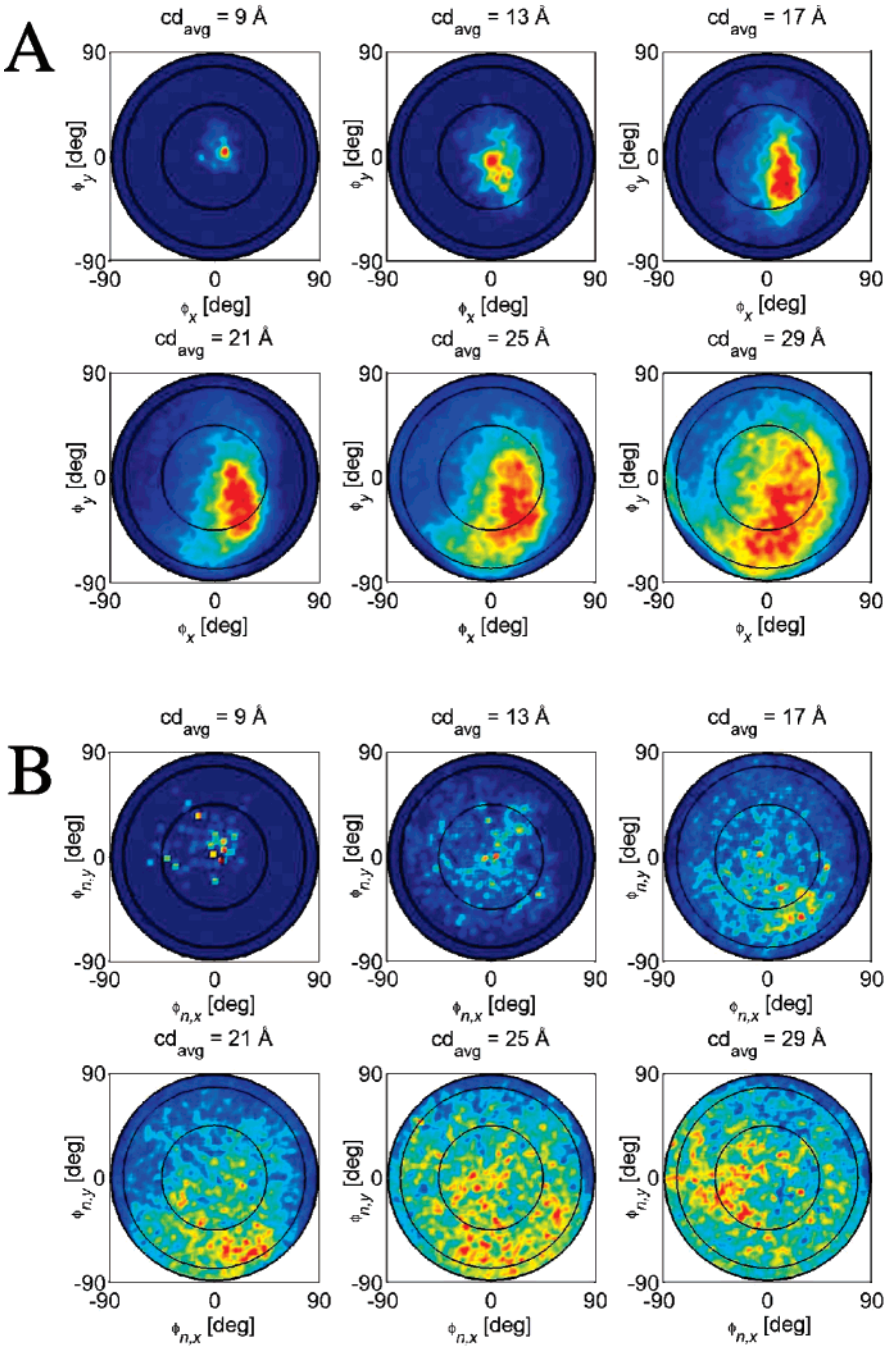

Figure 5. The computed occupancy maps after 10 000 trajectories as a function of the average contact distance cd_{avg}: (A) for the positional coordinates ϕ_x and ϕ_y and (B) for the orientational coordinates $\phi_{n,x}$ and $\phi_{n,y}$. Red means high occupancy and blue low occupancy.

the proteins form the encounter complex (see below). The splitting into the translational and rotational parts allows the separate computation of the entropy maps with respect to the positional and orientational coordinates. The computed values are stored in a further set of matrices, representing the entropy landscape. These matrices have the same grid sizes as the occupancy and energy matrices (see above). The entropy loss is computed separately for all grid nodes of the spatial and angular space which together represent the entropy landscape.

Since the proteins are simulated as rigid bodies we do not account for the internal entropy loss of the proteins concerning vibrational modes and side chain conformations. However, it can be assumed that this entropy contribution only becomes important for protein−protein separations of a few Å, i.e., in a regime, where the protein motions are not diffusion-limited anymore. Also, the entropy of the solvent is not considered here. Again, this contribution can be assumed to be limited to small protein−protein distances.

If the process of protein encounter is considered as a step-by-step process on a molecular level, it is clear that a protein at a certain position and with a certain orientation cannot explore the full configuration space within a Brownian dynamics time step. Therefore, we are interested in calculating the position and orientation dependent (local) entropy loss, rather than the total (global) entropy loss during the

association process. For the computation of the local entropy loss it seems reasonable to take into account the occupancies of all those configurations (positions and orientations), which are reachable from the particular position and orientation within one Brownian dynamics time step, i.e., which are within its *accessible volume* of the configuration space. The size of the accessible spatial and angular volumes V and $\mathscr{Y}$ should be in the range of the mean positional and orientational displacements during a Brownian dynamics time step. Since the protein motions are isotropic on average, we chose V as a sphere around the position of the protein and $\mathscr{Y}$ as a sphere around its orientation. The corresponding radii are $\Delta\rho$ and $\Delta\omega$, respectively.

In general, the entropy of a system with N states is given by

$$S = -k_B \sum_{n=1}^{N} P_n \ln P_n \tag{5}$$

where the P_n are the probabilities for each state n. In the case where all states are equally probable, i.e., $P_n = 1/N$ for all n, this is the famous formula $S = k_B \ln N$. To compute the local entropy loss at the given position and orientation from the occupancy of the positions and orientations (the states) within the accessible volumes V and $\mathscr{Y}$, we applied the basic formula, eq 5, by interpreting the occupancy landscape as a probability distribution. This configuration dependent entropy value is then compared to the entropy with a constant, isotropic probability distribution, which is the reference state if the proteins are far apart. Note that the entropy loss is computed separately for all grid nodes of the positional and orientational configuration space, which finally represent the entropy landscape. The translational entropy loss relative to the reference state at a given position is calculated in the following way: If the occupancy of a given state n in the volume V is F_n and the total occupancy of all states within V is F_{tot}, then the probability for state n is $P_n = F_n/F_{tot}$. According to eq 5 the entropy is

$$S_{tr} = -k_B \sum_{n=1}^{N} P_n \ln P_n = k_B \left(\ln F_{tot} - \frac{1}{F_{tot}} \sum_{n=1}^{N} F_n \ln F_n \right) \tag{6}$$

where N is the number of states within V, i.e., the number of grid points in this volume. Relative to the isotropic reference state with $P_n = 1/N$, the translational entropy loss during the process of protein−protein encounter is

$$\Delta S_{tr} = S_{tr} - k_B \ln N \tag{7}$$

The rotational entropy loss is computed analogously. It should be noted that the numerical value of the entropy as calculated by the above procedure is sensitive to the resolution of the occupancy maps (i.e. the number of states within the volumes V and $\mathscr{Y}$). We tested grid sizes between 21 and 201 nodes. The computed entropy curves (not shown) are quite similar, and only the smallest resolution results in too large values due to the decreasing statistics.

According to the formulas for the Brownian motion of two interacting proteins, eqs 2 and 3, the average translational and rotational displacement of barnase and barstar during a time step of $\Delta t = 1.0$ ps is $\Delta r = \sqrt{6D\Delta t} \simeq 0.4$ Å and $\Delta w_i = \sqrt{6D_{iR}\Delta t} \simeq 1.3°$ at a temperature of 300 K. These are, however, minimum values for $\Delta\rho$ and $\Delta\omega$ (the radii of V and $\mathscr{Y}$), since the displacements can be larger in a combined translational and rotational motion. Additionally, the average translational displacement is much larger when measured with respect to the contact distance, instead of the center-to-center distance. Some further, more technical aspects have to be taken into account for the choice of $\Delta\rho$ and $\Delta\omega$: On one hand, they should not be smaller than the resolution of the occupancy maps to ensure that the number of states within the volume of the configuration space $V\mathscr{Y}$ is not too small. On the other hand, $\Delta\rho$ and $\Delta\omega$ should be in the range of the average displacement in a BD time step to keep the idea of the local entropy definition.

Free Energy Landscape. In the sections above we described the computation of the energy and entropy landscapes. The contributions of electrostatic and desolvation energies as well as the translational and rotational entropy loss are stored in matrices, which represent the 6-dimensional configuration space. These matrices have the same grid sizes as the occupancy maps, and they are computed separately for the positional and orientational coordinates. In the energy matrices the minimum values for the given position/orientation are stored, finally allowing to identify the minimum free energy path. The translational and rotational entropy maps are calculated by interpreting the occupancy maps as probability distributions, i.e., by the restriction of the degrees of motional freedom.

With the energy and entropy contributions as functions of the translational and rotational coordinates, the free energy landscape of the encounter process is given by the sum of the electrostatic energy, the desolvation energy, and the translational/rotational entropy:

$$\Delta G = \Delta E_{el} + \Delta G_{ds} - T\Delta S_{tr},\ \Delta S_{tr} = \Delta S_{trans} + \Delta S_{rot} \tag{8}$$

Note that the free energy is calculated for each grid node of the spatial and angular space, representing the free energy landscape of the protein−protein encounter.

The reaction path is defined as the path along the minima of the free energy landscape. In Figure 6 the electrostatic interaction energy (A), the (negative) translational and rotational entropy loss (B), the desolvation energy (C), and the free energy (D) of protein−protein encounter along the reaction path is displayed for $\Delta\rho$ and $\Delta\omega$ varying from 2 to 5 Å and 2 to 5°, respectively. $\Delta\rho$ and $\Delta\omega$ denote the sizes of accessible volumes, i.e., the regions which are reachable from a certain position and orientation within one Brownian dynamics time step (see above). The entropic contribution is only weakly dependent on the volumes V and $\mathscr{Y}$ (Figure 6B). In all further simulations we used $\Delta\rho = 3$ Å and $\Delta\omega = 3°$. These values were kept constant also for varying temperature, since the time step in these simulations was adjusted to the average displacement at $T = 300$ K.

Results

The trajectories from BD simulations of protein−protein encounter were analyzed for the model system barnase−

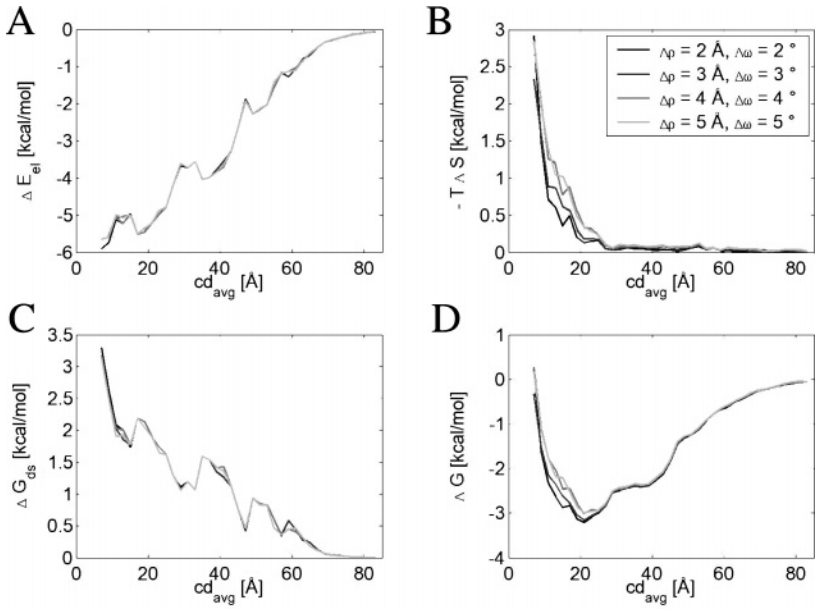

Figure 6. Contributions to the encounter free energy for varying volumes V and $\mathscr{V}$ in the entropy definition, i.e., for varying $\Delta\rho$ and $\Delta\omega$, as a function of the average contact distance: (A) the electrostatic energy, (B) the desolvation energy, and (C) the translational and rotational entropy. In (D) the free energy is displayed, the characteristic minimum at 21 Å denotes the position of the encounter complex.

barstar in order to clarify the nature of the encounter state. As explained in the previous chapter, the occupancy of the conformation space for the protein−protein encounter of barnase and barstar was computed on grids representing the 6-dimensional configuration space: one set of 3-dimensional matrices for the translational coordinates, one set for the rotational coordinates $\phi_{n,x}$ and $\phi_{n,y}$ plus the distance coordinate, and another matrix for ω_n and the distance coordinate. In Figure 5 the occupancy is displayed as a function of the average contact distance cd_{avg}, ranging from 9 to 29 Å with a step size of 2 Å (here only every second image is shown). While the average contact distance between the reaction patches of barnase and barstar in the crystal structure of the bound state is 4.0 Å, the closest configuration found in any trajectory is at 7 Å. It is to be expected that the particle density at such close distances is underestimated due to the neglect of adhesive interactions such as hydrophobic forces. The results shown in Figures 5−8 refer to the trajectory analysis of barnase and barstar at 300 K and 50 mM. In Figure 9 the dependence on ionic strength and temperature is investigated.

Figure 5A displays the occupancy as a function of the positional coordinates ϕ_x and ϕ_y (i.e. of θ and ϕ) for increasing cd_{avg}. Every disk shows the occupancy on the upper half (for $\theta < \pi$) of a shell with a thickness of 2 Å, projected on the plane perpendicular to the center-to-center vector of the proteins in the bound state (see Figure 4A). The black circles denote $\theta = 30°$, 60°, and 90°. For small cd_{avg} up to 17 Å, the occupied positions are quite close to the position of the proteins in the bound state, while the occupancy cloud is spread wider for increasing distance. At larger distances this cloud also becomes more anisotropic with respect to the angle ϕ, and the maximum of the occupancy cloud moves downward. For even larger contact distances ($cd_{avg} > 35$ Å) the maximum of the occupancy cloud is found in the leftmost part of the maps as indicated already in the lower right map ($cd_{avg} = 29$ Å). This means that the encounter is steered along a certain pathway towards the guanine binding loop of barnase. This nonisotropic encounter of the proteins can be understood from the charge distributions of barnase and barstar. The occupancy maps for the lower half spheres are not shown, since they are not significantly populated for distances shorter than $cd_{avg} \sim 40$ Å.

In Figure 5B the occupancy is shown as a function of the orientational coordinates $\phi_{n,x}$ and $\phi_{n,y}$ (i.e. of θ_n and ϕ_n), again for the upper half sphere ($\theta_n < \pi$). These maps are constructed similarly to the maps described above for the translational coordinates (see Figure 4B). For small cd_{avg} the orientation of the proteins is close to that in the bound state. With increasing contact distance the occupancy cloud is getting larger but also anisotropic with respect to the interface normal of protein 1. Again, the center of the occupancy cloud is displayed downward. Together with the interpretation of Figure 5A (an encounter of protein 2 from the left side is favorable) this means that already at a contact distance of 25 Å the binding interfaces of the proteins are preoriented and pointing roughly toward the binding interface of the other protein. The occupancy as a function of the third orientational coordinate ω_n (not shown here) has a funnel-like shape as well, converging toward $\omega_n = 0°$ for small contact distances, i.e., toward the configuration of the proteins in the bound state.

From these occupancy maps the entropy landscape was computed by the use of the above-defined local entropy function. Together with the electrostatic and desolvation energy, the free energy landscape is calculated as the sum of all terms. As mentioned above, the reaction pathway of the protein−protein encounter is given as the path along the minima in the free energy landscape. Figure 6 displays the

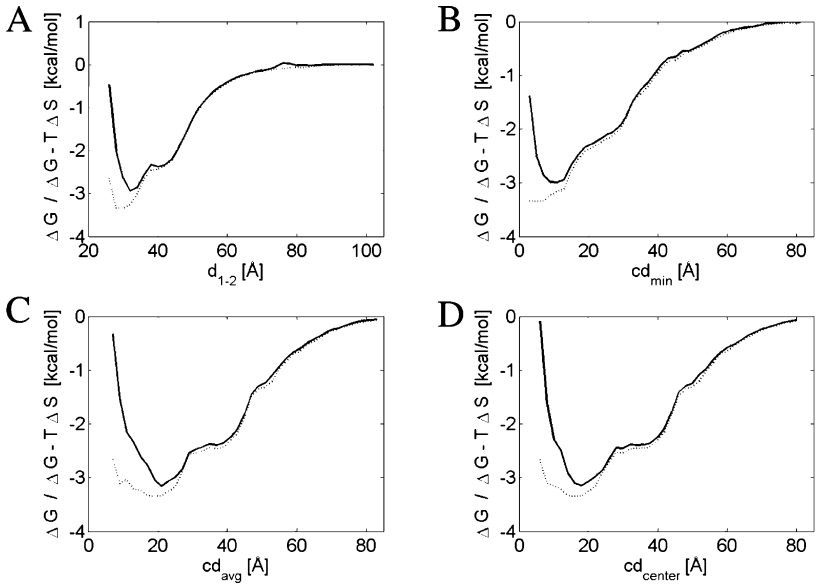

Figure 7. The profile of the encounter free energy as a function of the 4 different definitions of the distance axis (see Figure 3): (A) the center-to-center distance d_{1-2}, (B) the minimum contact distance cd_{min}, (C) the average contact distance cd_{avg}, and (D) the distance between the geometric centers of the reaction patches cd_{center}.

energy profiles along the reaction path as functions of the average contact distance for varying volumes V and $\mathscr{J}$ in the entropy calculation, $\Delta\rho$ and $\Delta\omega$ ranging from 2 to 5 Å and 2 to 5°, respectively. The electrostatic interaction energy (A) strongly decreases as the proteins get closer down to −6 kcal/mol, while the desolvation energy (C) is moderately increasing up to 3 kcal/mol. The (negative) entropic contribution (B) is negligible for $cd_{avg} > 25$ Å but increases rapidly as the proteins approach each other. At a distance of 8 Å the entropy loss amounts to ∼2.5 kcal/mol. The encounter free energy (D) decreases up to a distance of about 20 Å due to the favorable electrostatic interaction and increases again for smaller distances because of the desolvation energy and the entropy loss. The characteristic minimum between appears as a useful definition for the position of the encounter complex. In the following, we refer to 'minimum' when talking about the one-dimensional free energy profiles for protein−protein encounter. The term 'encounter complex' is used to discuss the conformations corresponding to these minima.

Having reached the encounter state, the proteins may spend some time in this free energy minimum in order to rearrange their interfaces including the orientation of the side chains. Moving the proteins closer than the encounter complex involves on one hand the (for barnase and barstar) unfavorable desolvation of the proteins and finally the loss of approximately all degrees of freedom of one of the proteins but on the other hand also the favorable formation of hydrophobic contacts, hydrogen bonds, and salt bridges. The entropy loss at the position of the minimum in Figure 6D is still very small (∼0.2 kcal/mol). This means that almost all the entropy is lost during the binding process from the encounter to the bound complex. The variation of the volumes V and $\mathscr{J}$ for the computation of the entropy leads to an increase of the entropy loss for increasing $\Delta\rho$ and $\Delta\omega$. This increase is only moderate leading to a slight shift of ΔG for $cd_{avg} \leq 20$ Å. For the subsequent simulations we chose $\Delta\rho = 3$ Å and $\Delta\omega = 3°$. These values are in the range of the BD time steps and of the resolution of the occupancy maps.

In Figure 7 (solid lines) the profile of the encounter free energy along the reaction pathway is shown for different definitions of the distance axis (see Figure 3): (A) the center-to-center distance d_{1-2}, (B) the minimum contact distance cd_{min}, (C) the average contact distance cd_{avg}, and (D) the distance between the centers of the reaction patches cd_{center}. All four graphs show the characteristic minimum in the profile denoting the position of the encounter state. The free energy values of the minimum conformations are quite close to each other ranging from −2.9 to −3.2 kcal/mol. The differences arise from the varying discretizations of the configuration space. The dotted lines show the sum of the electrostatic and desolvation energies, i.e., the free energy without the entropy loss. In all graphs the sum of the electrostatic and desolvation energies at the position of the minimum is lower than the free energy, while the position of the minimum is shifted toward smaller distances. These differences emphasize the importance of the entropy contribution for the definition of the encounter state.

To illustrate the effect of the four different definitions of the distance axis, representative pdb-structures of the encounter complex were constructed from the positions and orientations of the corresponding minimum free energy conformations. In Figure 8 they are shown for comparison: barnase is displayed as a gray van der Waals surface, barstar is schematically shown as a tube with the atoms of the reaction patch as spheres. For the bound structure (in black) only the reaction atoms are displayed. The encounter complex conformation of barstar is colored in blue for the minimum position in the free energy profile with the distance axis d_{1-2} (Figure 7A), in red for cd_{min} (Figure 7B), in green for cd_{avg} (Figure 7C), and in purple for cd_{center} (Figure 7D). The positions of the encounter complex for the different definitions of the distance axis and the corresponding RMSDs to

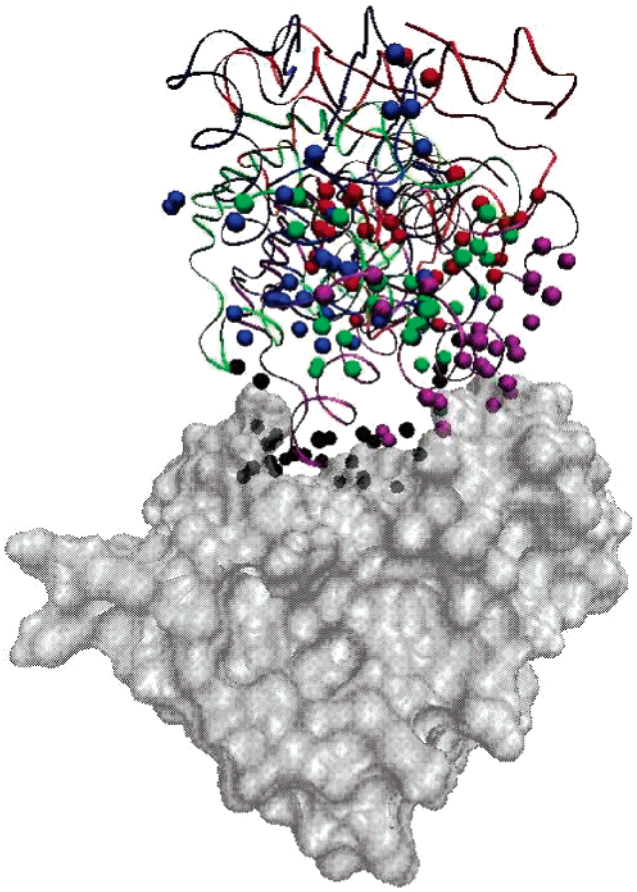

Figure 8. Comparison of representative conformations of the encounter complex structures, as defined by the minima in the encounter free energy profiles in Figure 7. Barnase is displayed as a gray van der Waals-surface, barstar is schematically shown as a tube with the reaction atoms as spheres. For the bound structure (in black) only the reaction atoms are displayed. The encounter complex conformation of barstar is colored in blue for the minimum position in the free energy profile with the distance axis d_{1-2} (Figure 7A), in red for cd_{min} (Figure 7B), in green for cd_{avg} (Figure 7C), and in purple for cd_{center} (Figure 7D).

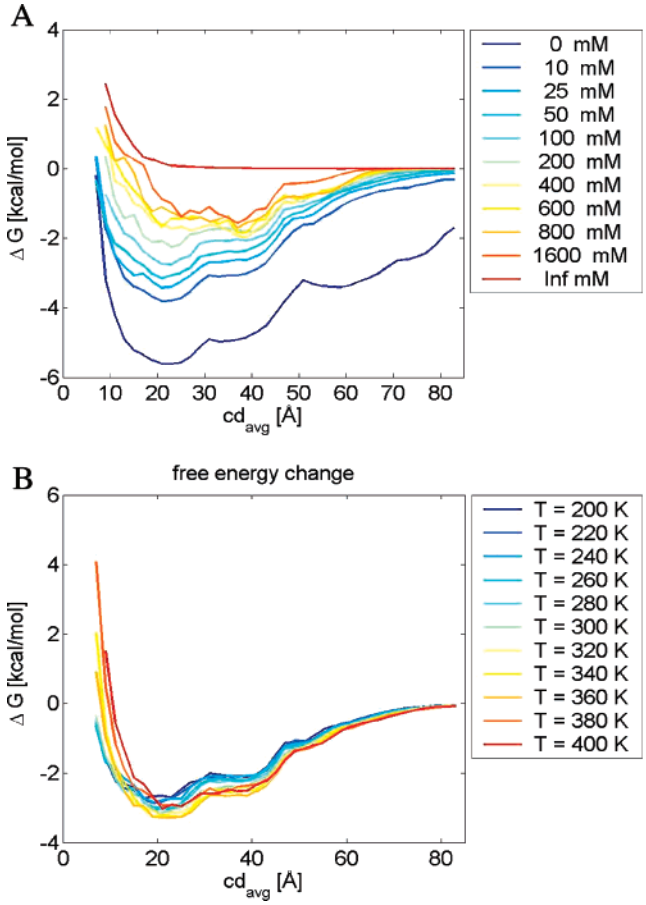

Figure 9. The encounter free energy profiles (A) for varying ionic strength at a constant temperature of 300 K and (B) for varying temperature at a constant ionic strength of 50 mM.

the bound configuration are as follows: $d_{1-2} = 32$ Å (RMSD = 16.4 Å), $cd_{min} = 11$ Å (RMSD = 18.6 Å), $cd_{avg} = 21$ Å (RMSD = 18.2 Å), and $cd_{center} = 18$ Å (RMSD = 18.7 Å). The pairwise distance between the protein centers in the 4 encounter complex configurations is 7.1 ± 5.2 Å. The pairwise RMSD, however, has a large spread: 14.4 ± 7.2 Å. This indicates that the translational degrees of freedom are much more restricted in the encounter complex than the rotational degrees of freedom. The occupancy maps in Figure 5 show the same behavior. The occupancy as a function of the positional coordinates at $cd_{avg} = 21$ Å is confined to a rather small area (lower left image of Figure 5A), while the occupancy cloud for the orientational coordinates at the same distance is still quite large (lower left image of Figure 5B).

Satisfied by these results we added two series of simulations in order to investigate the dependence of the free energy profile, in particular the position of the minimum, on the ionic strength and on the temperature of the solvent. For the analysis of the ionic strength dependence, the temperature was set to 300 K, while the ionic strength was varied in logarithmic steps from 0 mM up to 1600 mM. One simulation was performed where the electrostatic and desolvation forces were switched off in order to model infinite ionic strength. The results of this study are shown in Figure 9A. The depth of the free energy minimum decreases continuously with increasing ionic strength due to increasing shielding of the electrostatic interactions. The entropy contribution (not shown separately) is almost independent of the ionic strength. The distance of the minimum is approximately 22 Å for low and moderate ionic strength, ranging from 0 mM up to 200 mM. [These values are computed by a parabolic fit of the free energy profile around the minimum position.] At even higher values this minimum disappears involving a rapid jump of the encounter complex position to ≥ 38 Å. At infinite ionic strength the free energy profile has no minimum since the entropy loss is the only contribution. Further sets of simulations were performed with the same values for temperature and ionic strength but using the linear instead of the full Poisson−Boltzmann equations (PBE) for the computation of the electrostatic grids. Although the depth of the energy minimum is more shallow in the simulations where the full PBE were used, the shapes of the free energy profiles (not shown) for varying ionic strength are quite similar to the profiles displayed in Figure 9A. For example, the minimum free energy value at 50 mM and 300 K is −3.2 kcal/mol when the full PBE were used and −4.4 for the linear PBE. This effect is most relevant for values of

the ionic strength ≤ 200 mM, but also the curves at larger values are slightly shifted.

In the simulations where the full PBE were used, the solvent dielectric ϵ_s was varied according to its temperature dependency, while in the other simulations (where the linear PBE were used) it was set to the value of 78. As mentioned above, in all simulations the phase transitions of water at 273 and 373 K and an eventual unfolding of the proteins were omitted. The energy and entropy profiles (not shown) for the latter set of simulations (using the linear PBE at constant ϵ_s) reflect the expected behavior: For increasing temperature the electrostatic and desolvation energies decrease slightly. This is accompanied by a small increase of the entropy. Both effects can be understood by the increasing mobility of the proteins at increasing temperature. The overall shape of the free energy profiles is quite similar, while the depth of the potential well decreases moderately with increasing temperature.

The effect of the varying solvent dielectric in the former set of simulations (using the full PBE at temperature dependent ϵ_s) is somehow surprising. It can be explained by the better ordering of the water molecules at low temperature and by the decreasing shielding of water at increasing temperature. Therefore, the electrostatic and desolvation energies increase for increasing temperature. Again, the entropy slightly increases for increasing temperature. The effects of increasing electrostatics, desolvation, and entropy almost cancel each other. Thus, the free energy profiles (Figure 9B) seem to be somehow independent of the temperature. For increasing temperature the distance of the minimum systematically increases from 19.7 Å up to 23.6 Å.

Discussion

By analyzing Brownian dynamics trajectories of barnase and barstar we gained additional insight into the mechanism of protein−protein encounter, information that is complementary to the computation of their association rates. From the detailed evaluation of the trajectories we can learn about the encounter pathways as well as the degree of preorientation of the proteins with respect to their reaction patches. Furthermore, by the definition of a local entropy function, we are able to compute the free energy landscape as the sum of the electrostatic and desolvation energies plus the contribution from the translational and rotational entropy. In the free energy profile along the reaction pathway, defined as the path along the minima of the free energy function, a minimum at small protein−protein separations shows up. Having reached this minimum position after the initial free diffusion, the proteins may spend some time there in order to rearrange their interfaces for the formation of the bound complex. This characteristic minimum is therefore a suitable definition for the encounter complex, the intermediate state before the final complexation of the proteins. The positions of the encounter complex, computed for 4 different definitions of the distance axis, are close to each other, while the orientational degrees of freedom at the encounter state are not much restricted. This behavior is reflected in the occupancy maps in Figure 5. It underlines that the encounter state as defined by the minimum of free energy is not a single conformation like the bound state but rather a cloud of conformations with almost the same, minimal free energy.

It is often assumed that protein−protein complexes with strong electrostatic steering show a significant degree of preorientation of both binding partners. Our analysis suggests that this preorientation mostly affects the translational degrees of freedom, while the rotational preorientation takes place during the desolvation phase (between (3) and (4) in Figure 1). In another simulation at 50 mM and 300 K with an exchange of barnase and barstar (the motion of barnase is computed relative to barstar) we found that for barstar the translational degrees of freedom are more restricted than for barnase. Because of the stochastic nature of BD trajectories we did not consider it meaningful to distinguish between 'successful' and 'nonsuccessful' trajectories. At any point of the occupancy maps, an individual trajectory may proceed into any arbitrary direction. On the other hand, the behavior of ensembles of trajectories is determined by the free energy funnel computed from the occupancy maps. It should be feasible to compute time scales for transitions between any two points by a Kramers-like approach. Such analysis will be presented in a subsequent manuscript.

One may certainly ask whether BD simulations are a suitable tool for defining the location of the encounter complex considering that the encounter complex is denoted as end-point of the diffusional encounter. As mentioned in the Introduction, typical BD simulations of protein−protein encounter make simplifications which become important when the proteins approach the bound conformation, at protein−protein separations shorter than ∼10 Å. In our studies we found the position of the free energy minimum at an average contact distance of about 20 Å. For the system studied here, the region of the encounter complex is therefore in the regime where the BD simulation would still be valid.

The distance of the minimum position ($cd_{avg} \simeq 21$ Å) corresponds to a minimum contact distance of about 11 Å (see Figure 7B). This distance is larger than the value of 7.5 Å which was suggested by Gabdoulline and Wade by comparing the computed association rates with experimental data for the 3-contacts criterion[3] and is significantly larger than the 6 Å which was used for the criterion of 2 independent contacts between hydrogen-bond donors and acceptors.[1] These authors denoted this position as the *diffusional encounter complex*, the end point of diffusional encounter at which the proteins are committed to form a bound complex. It can be assumed to be located just after the transition state (see Figure 1). Since this position is closer to the bound complex than the encounter complex as defined in this work by the minimum of free energy, the association rate at the position of the diffusional encounter complex is significantly reduced compared to the rate at the free energy minimum. We computed the association rates in the simulation at 300 K and 50 mM ionic strength as 3.44×10^8 M^{-1} s^{-1} for 2 independent donor−acceptor contacts shorter than 6.0 Å (i.e. for the diffusional encounter complex) and as 4.07×10^9 M^{-1} s^{-1} for 1 donor−acceptor contact shorter than 10 Å (i.e. for the encounter complex). For comparison, the experimental value is 2.86×10^8 M^{-1} s^{-1}.[9] The computed

values are expected to overestimate the measured rates by about 30% due to the neglect of the hydrodynamic interactions.[1] The position of the diffusional encounter complex defined by Gabdoulline and Wade therefore yields computed rates that are comparable to experimental values. On the other hand, the state of minimal free energy characterized in this work should be useful for interfacing BD simulations for protein−protein encounter with atomistic MD simulations to study the details of the final steps during the binding process. The free energy minimum should characterize a short-lifetime intermediate state that may eventually become accessible to experimental observations. As a note of caution one should add that many interacting protein−protein pairs do not show electrostatic complementarity.[1] It remains to be seen whether these systems also show a small free energy minimum before the onset of freezing rotational and translational degrees of freedom.

A more detailed analysis of the reaction pathway and the occupancy maps in Figure 5 shows that barstar is indeed steered toward the guanine binding loop of barnase in consistency with the results of Gabdoulline and Wade[10] and Camacho et al.[7] Along the encounter process the proteins are guided by a free energy well: At large distances, barstar moves toward Lys66 (in Figure 5A from the left) and then turns to the center of the barnase reaction patch, and, at $cd_{avg} \sim$ 25 Å, barstar is steered toward the guanine binding loop of barnase converging to its bound conformation. This encounter behavior partly reproduces the result of the Boltzmann factor calculation of Gabdoulline and Wade (Figure 5(a) in ref 1). The middle and back part of their probability distribution is similar to the described pathway. The front part, however, was found to be energetically unfavorable.

The free energy profiles were evaluated for varying ionic strength and temperature. The variation of the ionic strength has a large impact on the profiles, both for the depth and for the position of the minimum. The dominating effect is the increasing shielding of the electrostatic interactions due to the increasing ionic strength of the solvent which results in a destabilization of the encounter state for values of the ionic strength ≥ 400 mM. The variation of the temperature of the solvent, in contrast, has only a small effect on the free energy profiles: For increasing temperature the electrostatic and desolvation energies as well as the entropy increase in a manner that their effects on the free energy profile almost cancel out each other. The distance of the minimum increases continuously by about 4 Å for increasing temperature. From these studies we can conclude that for low and moderate ionic strength the position of the encounter complex is slightly dependent on the temperature but almost constant for varying ionic strength. At high ionic strength, however, the encounter complex is destabilized involving a jump of the position of minimum free energy. The depth of the minimum varies about 4 kcal/mol between 10 and 1600 mM ionic strength and about 0.6 kcal/mol in the temperature range from 200 to 400 K. The stability of the encounter complex is therefore mostly affected by varying ionic strength compared to variations in temperature. For protein systems with weaker electrostatic interactions we can therefore expect that the free energy well at the encounter complex is flatter. The contribution of the translational and rotational entropy is expected to depend strongly on the steepness of the encounter funnel.

Studies of this kind may provide a basis for theoretical predictions of FRET signals for protein−protein interaction. Also, they may shed light on the interesting question why protein surfaces have been formed by evolution into what we find today. In the future we want to apply our simulation method to more complex systems such as the encounter of cytochrome *c* and cytochrome *c* oxidase.[26] The focus will be more on the analysis of the association and dissociation pathways (ongoing work) and also on time-dependent aspects of association such as the lifetime of the observed free energy minimum.

Acknowledgment. We gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft via the Center for Bioinformatics in Saarbrücken. We thank Rebecca C. Wade for the provision of their SDA program and valuable comments on the manuscript. A. Spaar thanks Dagmar Flöck for helpful advice in the beginning of this project.

References

(1) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139−1155.

(2) Schreiber, G. *Curr. Opin. Struct. Biol.* **2002**, *12*, 41−47.

(3) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Recognit.* **1999**, *12*, 226−234.

(4) Gabdoulline, R. R.; Wade, R. C. *Methods* **1998**, *14*, 329−341.

(5) Janin, J. *Proteins* **1997**, *28*, 153−161.

(6) Vijayakumar, M.; Schreiber, K. Y. W. G.; Fersht, A. R.; Zhou, A. S. H. Z. *J. Mol. Biol.* **1998**, *278*, 1015−1024.

(7) Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. *Biophys. J.* **1999**, *76*, 1166−1178.

(8) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878−8889.

(9) Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995**, *248*, 478−486.

(10) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917−1929.

(11) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405−414.

(12) Frisch, C.; Fersht, A.; Schreiber, G. *J. Mol. Biol.* **2001**, *308*, 69−77.

(13) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(14) QUANTA, *Molecular modelling software package;* Molecular Simulations, Inc.: San Diego, CA, 1992.

(15) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149−162.

(16) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Comm.* **1991**, *62*, 187−197.

(17) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(18) Archer, D. G.; Wang, P. *J. Phys. Chem. Ref. Data* **1990**, *19*, 371.

(19) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868−3878.

(20) Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. *J. Phys. Chem.* **1987**, *91*, 5991−5998.

(21) Northrup, S. H. *MacroDox v.2.0.2: Software for the Prediction of Macromolecular Interaction;* Tennessee Technological University: Cookeville, TN, 1995.

(22) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352−1360.

(23) Jones, R. A. L. *Soft Condensed Matter;* Oxford University Press: Oxford, 2002.

(24) Antosiewicz, J.; Briggs, J. M.; McCammon, J. A. *Eur. Biophys. J.* **1996**, *24*, 137−141.

(25) Gorba, C.; Geyer, T.; Helms, V. *J. Chem. Phys.* **2004**, *121*, 457−464.

(26) Flöck, D.; Helms, V. *Biophys. J.* **2004**, *87*, 65−74.

(27) Ben-Tal, N.; Honig, B.; Bagdassarian, C. K.; Ben-Shaul, A. *Biophys. J.* **2000**, *79*, 1180−1187.

(28) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33−38.

CT050036N