Biophysical Journal, Volume 99

Supporting Material

Title: Absolute Protein-Protein Association Rate Constants From Flexible, Coarse-Grained Brownian Dynamics Simulations: The Role of Inter-Molecular Hydrodynamic Interactions in Barnase-Barstar Association

Authors: Tamara Frembgen-Kesner and Adrian H Elcock

METHODS

The structure and energetics of the pseudoatomic model

Generation of the pseudoatomic models began with all-atom representations of wild-type barnase and barstar taken from the crystal structure of the barnase-barstar complex reported by Buckle *et al.* (1; pdb code: 1BRS). In addition to wild-type barnase, three charge-mutants (E60A, R87A, and R59A) were selected for study on the basis of their wide-ranging effects on the association rate constant of the complex: the E60A mutation results in an approximately 5-fold increase over the wild-type association rate, while the R59A mutation results in an approximately 10-fold decrease (2,3). These mutants were modeled by removing the sidechain atoms beyond the Cβ atom of the mutated residue. To determine effective charges for the pseudoatomic models, electrostatic potentials around the corresponding all-atom models were first solved using the adaptive Poisson-Boltzmann solver (APBS) program (4). For these calculations hydrogen atoms were first added to each protein structure using the APBS utility PDBTOPQR (5,6); partial charges and radii were then assigned to atoms from the PARSE parameter set (7). The internal dielectric of the protein was set to 12.0, the solvent dielectric was set to 78.4, and the ionic strength was set to the corresponding experimental value, 50 mM. (For results with ionic strength set to 500 mM, see Additional Results.) The nonlinear PB equation was then solved on a 3D 129 × 129 × 129 grid of spacing 1 Å.

Pseudoatomic models were constructed by applying the *q_pdb* utility from the Situs suite of programs (8) to barnase and barstar separately; this utility fills the internal volume of a protein with a user-specified number of spheres (pseudoatoms) that are placed and sized to best replicate the protein's electron density envelope. The number of pseudoatoms applied to each protein was determined on the basis of the resulting model's ability to reproduce the electrostatic potential of its corresponding all-atom model. Specifically, the effective charge method (ECM) program (9) was used to derive effective charges for the placed pseudoatoms that, when expressed as a sum of Debye-Hückel potentials, best reproduced the all-atom potential in a 3 Å-thick layer around the protein structure. As expected, the greater the number of pseudoatoms (and, therefore, effective charges) in the model, the greater the ability to

accurately reproduce the computed electrostatic potential of the all-atom model. Empirically, we determined that a ratio of 3 pseudoatoms for every 10 residues was sufficient to give an average error of less than 10% in the computed potential (Fig. S1); this criterion also produced pseudoatomic models that reflect reasonably well the overall shapes of the proteins (see Fig. 1 of the main text.) The radii of the pseudoatoms generated by the *q_pdb* utility ranged from ~4.0 to 5.4 Å. Effective charges were derived separately for each of the barnase mutants; in all cases, however, the pseudoatom *positions* were identical with those of the wild-type barnase pseudoatomic model.

To maintain each model's shape during simulation, nearby pseudoatom pairs were joined to each other via flexible harmonic bonds, with the energy expressed by:

$$E_{bond} = \sum_{bonds} K_r (r - r_0)^2 \quad (1)$$

where E_{bond} is the total bonded energy of the model, K_r is the bond force constant (arbitrarily set here to 20 kcal/mol/Å^2), and r and r_0 are the current and 'native' bond lengths, respectively. The cutoff distance for deciding which pseudoatoms would be considered 'bonded' was adjusted so that, on average, three bonds per pseudoatom were formed. For barnase, pseudoatoms within 10.8 Å of each other were connected via bonds; for barstar, the cutoff was 11.0 Å. In addition to visual verification that the models retained their shape, snapshots were taken at 1 ns intervals over the course of separate 10 μs simulations of barnase and barstar. Each snapshot was subsequently superimposed upon the initial structure and the root mean squared deviation (rmsd) of all pseudoatoms was calculated. The averaged rmsd for barnase was 0.56 ± 0.16 Å; for barstar, it was 0.93 ± 0.61 Å.

The bond energy was the only *intra*molecular energy term included in the simulations. Intramolecular electrostatic interactions were omitted, while intermolecular interactions were calculated using the Debye-Hückel equation:

$$E_{ij}^{DH} = 332.08\, q_i q_j \exp(-\kappa r_{ij}) / \varepsilon r_{ij} \quad (2)$$

where q_i and q_j are the effective charges on pseudoatoms i and j; κ is the ionic strength; ε is the solvent dielectric constant; and r_{ij} is the center-to-center distance between the i'th and j'th pseudoatoms during simulation. No cutoff was applied to electrostatic interactions during the simulations.

Following previous work performed by one of us in the context of protein folding simulations (10), non-electrostatic intermolecular interactions were calculated using one of two models. First, those pseudoatom pairs not involved in close contacts in the complexed state were assigned a purely repulsive interaction:

$$E_{ij}^{nonnative} = \varepsilon \, (\sigma^{12} / r_{ij}^{12}) \qquad (3)$$

where ε is set to 0.1 kcal/mol, σ is set to 6.0 Å (1.0 Å less than the shortest distance between a 'Gō' pair [11] of pseudoatoms; see below) and r_{ij} is again the distance between i'th and j'th pseudoatoms during simulation. Second, those pseudoatom pairs involved in close contacts in the complexed state were assigned a *very* weakly favorable Lennard-Jones-type energy function:

$$E_{ij}^{native} = \varepsilon \, [\, 5 \, (\sigma_{ij}^{12} / r_{ij}^{12}) \, - \, 6 \, (\sigma_{ij}^{10} / r_{ij}^{10}) \,] \qquad (4)$$

where ε is the energy well depth of the potential (set to 0.1 kcal/mol), σ_{ij} is the distance between the two pseudoatoms in the structure of the native complex, and r_{ij} is the distance between the i'th and j'th pseudoatoms during simulation. Here, as in our previous work (12), a 'close contact' is defined as being two non-hydrogen atoms within 5.5 Å of each other in the all-atom model of the native complex; each such contact of atom pairs was converted into a 'contact' of pseudoatom pairs by assigning each atom to its nearest pseudoatom. Each pair of pseudoatoms involved in a contact was counted at most once; defined in this manner, we obtained a total of 34 native (Gō) contacts in the pseudoatomic model of the barnase-barstar complex. The well-depth, ε, assigned to these pseudoatom-pseudoatom interactions was deliberately kept small so that a double-counting of the driving force for complexation did not occur in the simulations: in line with previous work, the primary driving force for the barnase-barstar association in the simulations is assumed to be provided by attractive electrostatic interactions (2). In effect,

therefore, the Gō contacts serve only to provide a convenient measure of the extent to which the simulated barnase-barstar interaction during a trajectory resembles the native complex (see "Reaction criteria").

Brownian dynamics simulation protocol

As in our previous work (12), all Brownian dynamics simulations were conducted using in-house code implementing the Ermak-McCammon algorithm (13):

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \sum_j \mathbf{D}_{ij} \mathbf{F}_j \Delta t / k_B T + \mathbf{R}_i \quad (5)$$

where $\mathbf{r}_i(t)$ is the position vector of the i'th pseudoatom at time t; Δt is the timestep; $\mathbf{D}_{ij}$ is the i'th, j'th 3 × 3 submatrix of the diffusion tensor $\mathbf{D}$; $\mathbf{F}_j$ is the total force acting on the j'th pseudoatom; k_B is Boltzman's constant; T is the temperature in Kelvin (298 K) and $\mathbf{R}_i$ is the stochastic displacement of the i'th pseudoatom. To incorporate hydrodynamic interactions (HI), the elements of the diffusion tensor were calculated using the equations of Rotne & Prager and Yamakawa (14,15) with their extensions for pseudoatoms of unequal size (16,17):

$$\mathbf{D}_{ii} = (k_B T / 6\pi\eta_s\sigma_i)\, \mathbf{I} \quad (6)$$

$$\mathbf{D}_{ij} = (k_B T / 8\pi\eta_s)\,(1/r_{ij}) \times \{ [\, 1 + (\sigma_i^2+\sigma_j^2)/3r^2_{ij} \,]\, \mathbf{I} + [\, 1 - (\sigma_i^2+\sigma_j^2)/r^2_{ij} \,]\, (\mathbf{r}_{ij}\mathbf{r}_{ij}/r^2_{ij}) \} \quad \textit{for } r_{ij} \geq (\sigma_i+\sigma_j)$$

$$\underline{\text{or}} \quad \times [\, r_{ij}/(\sigma_i+\sigma_j) \,] \{ [\, 8/3 - 3r_{ij}/2(\sigma_i+\sigma_j) \,]\, \mathbf{I} + [\, r_{ij}/2(\sigma_i+\sigma_j) \,]\, (\mathbf{r}_{ij}\mathbf{r}_{ij}/r^2_{ij}) \} \quad \textit{for } r_{ij} < (\sigma_i+\sigma_j)$$

where $\mathbf{I}$ is a unit 3 × 3 matrix, σ_i and σ_j are the hydrodynamic radii of the i'th and j'th pseudoatoms, r_{ij} is the distance between the i'th and j'th pseudoatoms, and $\mathbf{r}_{ij}$ is the connecting vector. The radii assigned to the pseudoatoms by the *q_pdb* utility (~4.0 to 5.4 Å) were adjusted to reproduce the translational and rotational diffusion coefficients predicted for the corresponding all-atom protein models by the hydrodynamics program HYDROPRO (18,19). HYDROPRO's predictions for the diffusion coefficients of wild-type barnase, barstar and the barnase-barstar complex are reported in Table S1. A number of 10 μs BD simulations of each pseudoatomic model were performed, with the resulting translational and rotational diffusion

coefficients being calculated as described in our previous work (12). The radii were increased in 0.1 Å increments until satisfactory agreement with HYDROPRO's predictions was obtained; this was achieved when each of *q_pdb*'s assigned pseudoatomic radii were increased by 3.5 Å to ~7.5 to 8.9 Å. Notably, the hydrodynamic radii obtained from the independent simulations of barnase and barstar also accurately captured the HYDROPRO-predicted diffusion coefficients of the barnase-barstar complex (see Table S1).

We performed two basic types of BD simulations. In the first, we included intramolecular hydrodynamic interactions but excluded intermolecular hydrodynamic interactions. This requires setting all elements of the diffusion tensor that involve pseudoatoms on different molecules to zero; in effect, therefore, the complete 3N × 3N matrix of the two-protein system reduces to two smaller 3N × 3N matrices: one with N = 33 for barnase and another with N = 27 for barstar. In the second type of simulations we included all intramolecular and intermolecular hydrodynamic interactions; in such a case the entire 3N × 3N matrix (where N = 33 + 27 = 60) is required. In principle, a third set of simulations could have been performed in which all intramolecular and intermolecular hydrodynamic interactions are set to zero (this is known as a 'free draining' model). We did not pursue such simulations, however, because we know from our previous study that the translational and rotational diffusion coefficients of flexible protein models are drastically underestimated when all hydrodynamic interactions are omitted (12). Furthermore, the magnitude of the underestimation of the translational diffusion coefficient is worse by several-fold than it is for rotational diffusion coefficients: in general, therefore, experimental translational and rotational diffusion coefficients cannot be replicated simultaneously in free draining (no HI) simulations of flexible models.

As in our previous work (12), we followed the standard approach of computing the correlated random displacements required in the Ermak-McCammon algorithm by performing a Cholesky decomposition of the diffusion tensor. Routines for calculating the diffusion tensor and its Cholesky decomposition were modified from those given by Allen and Tildesley (20). In all BD simulations the timestep was set to 250 fs and the elements of the diffusion tensor and its decomposition were recomputed every 100 simulation steps (i.e. every 25 ps).

Calculation of association rate constants

As in previous all-atom BD studies (e.g., 21–25), association rate constants were calculated by analyzing the results of a series of BD simulations with the Northrup-Allison-McCammon method (21). In this method, each BD simulation begins with the two proteins positioned in random relative orientations at a center-to-center separation distance, b. The trajectory is continued until the two proteins either exceed a specified separation distance, c (where $c >> b$), or satisfy the reaction criteria used to define formation of a complex (see below). Following others who have simulated the barnase-barstar association reaction (22,24,25), we set b and c to 100 Å and 500 Å respectively. The reaction rate constant k is then derived using the fraction of all simulated trajectories, β, that satisfy the reaction criteria using:

$$k = k(b) \times \beta \,/\, [\, 1 - (1 - \beta) \times k(b)/k(c)] \qquad (7)$$

where $k(b)$ and $k(c)$ are, respectively, the rate constants for two spheres achieving relative separations of b and c; in the present case, both b and c are sufficiently large that these rate constants are given simply by the Smoluchowski result: $k(x) = 4\pi Dx$, where D is the relative translational diffusion coefficient, i.e. the sum of the two proteins' translational diffusion coefficients. HYDROPRO's estimate of the relative translational diffusion coefficient of barnase and barstar monomers is 2.672×10^{-2} Å^2/ps. The term β represents the number of simulations that satisfied the reaction criteria (see below).

In order to obtain statistically robust estimates of association rate constants it was necessary to compute a large number of trajectories: from 50,000 to 250,000 (see Table S2). The complete set of simulations required approximately 11 weeks running on ten 8-core intel CPUs.

Reaction criteria

The Gō-like model (11) used to describe non-electrostatic intermolecular interactions provides a natural way to identify successful reactions. The extent of the encounter, if any, is quantified throughout the simulations in terms of the structure (order) parameter 'Q' (see, e.g. refs. 26 & 27), which is defined as the number of intermolecular Gō contacts formed at a given point in the

simulation *divided by* the total number of contacts – 34 – that would be found in the fully complexed state; Q therefore varies between 0 and 1. During each simulation trajectory the highest value of Q achieved is recorded; β is then calculated by counting all trajectories in which the highest attained value of Q satisfies or exceeds a given threshold value, Q_{rxn}. It is to be noted that although the terminology used in our method is borrowed from the protein folding community, it is conceptually equivalent with the approach used in previous BD simulations of protein-protein association events (e.g., 22): in both cases, successful associations are defined by the formation of some fraction of the residue-residue interactions found in the structure of the native complex.

ADDITIONAL RESULTS

Additional simulations of the wild type barnase and barstar association were performed with the ionic strength set to 500 mM. For these simulations, the same pseudoatomic models employed in the 50 mM simulations were again used but with the effective charges recalculated to fit PB electrostatic potentials calculated at 500 mM; all other features of the simulations, including the choice of Q_{rxn}, were identical with those used in the 50 mM simulations. The experimental association rate constant (2) of wild type barnase and barstar is decreased by ~20-fold when the ionic strength of the solution is increased from 50 mM to 500 mM (i.e. from 2.86×10^8 M^{-1} s^{-1} to 0.14×10^8 M^{-1} s^{-1}). In simulations performed without intermolecular HI, we find that the association rate constant decreases by a factor of 2 (i.e. from from 2.77×10^8 M^{-1} s^{-1} to 1.36×10^8 M^{-1} s^{-1}); in simulations performed with HI, we again find that the association rate constant decreases by a factor of 2 (i.e. from 2.54×10^8 M^{-1} s^{-1} to 1.23×10^8 M^{-1} s^{-1}). Clearly, therefore, in terms of reproducing the effects of increasing the ionic strength of the solution, the simulations are in only qualitative agreement with experiment. One possible contributor to the poor quantitative agreement at 500 mM may be the reduced ability of the effective charges to reproduce the PB-computed electrostatic potential: for barnase, the error in the potential computed from the effective charges was 18.0% at 500 mM (vs. 6.5% at 50 mM), whereas for barstar, the error was 13.0% (vs. 3.0% at 50 mM). While it might be thought that more structurally detailed models would produce better results, it is important to note that an

underestimation of the screening effects of salt appears to be a consistent feature of the effective charge approach: in Gabdoulline and Wade's original study of the barnase-barstar system (22), for example, the computed rate constant for association in 500 mM was found to be only ~3-fold lower than that obtained at 50 mM. It is possible, therefore, that a quantitative reproduction of the effects of increasing ionic strength might only be achieved with a simulation model that explicitly includes the dissolved salt ions.

FIGURES AND TABLES

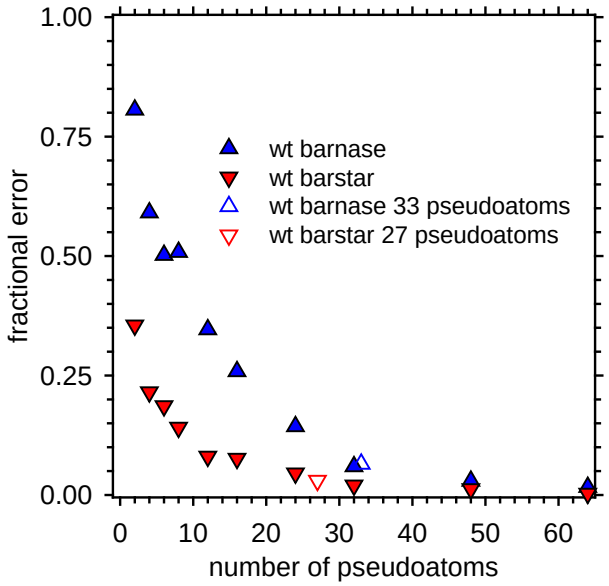

Figure S1 The error in the computed electrostatic potential due to the effective charges of the pseudoatomic models plotted versus the number of pseudoatoms in the model. Errors for barnase mutants are similar to those shown above: for 33-pseudoatom models the percent errors were 6.5%, 5.4%, 7.9%, and 8.7% for the wild-type, E60A, R87A and R59A mutants respectively.

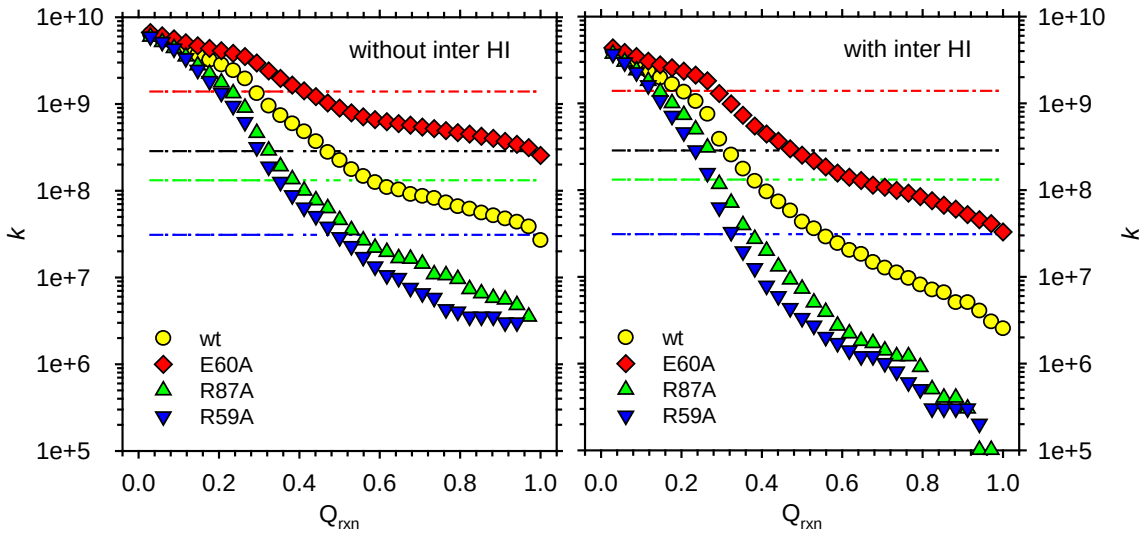

Figure S2 Association rate constants plotted v. Q_{rxn}. Horizontal dashed lines represent experimental values and are color-coordinated, with the exception of wild-type (yellow circles coupled with a black dashed line). The association rate constants decrease with increasing stringency (i.e. increasing Q_{rxn}) in all simulations. Interestingly, however, the rate of decrease is significantly greater in simulations with intermolecular HI.

Table S1 Translational and Rotational Diffusion Coefficients: Comparison of HYDROPRO's predictions to the results of BD simulations (units in 10^{-2} Å^2 ps^{-1} and 10^{-2} ns^{-1}, respectively)

Protein	# Res	Pseudo Atoms	HydroPro D_{trans}	Simulated D_{trans}	100% × Sim/HPro	HydroPro D_{rot}	Simulated D_{rot}	100% × Sim/HPro
Barstar	89	27	1.40	1.41	100.8%	3.27	3.47	106.1%
Barnase	110	33	1.28	1.31	102.4%	2.45	2.44	99.9%
Barnase-Barstar	199	60	1.04	1.08	103.2%	1.33	1.28	96.6%

Table S2 Numbers of trajectories used to compute association rate constants

	# Sims w/o Inter HI	β w/o Inter HI	# Sims w/ Inter HI	β w/ Inter HI
wt	50,000	0.011	50,000	0.010
E60A	50,000	0.041	50,000	0.039
R87A	100,000	0.0025	250,000	0.0028
R59A	100,000	0.0015	250,000	0.0013
wt – 500mM	50,000	0.0054	100,000	0.0049

REFERENCES

1. Buckle, A. M., G. Schreiber, and A. R. Fersht. 1994. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry*. 33:8878–8889.
2. Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol*. 3:427–431.
3. Gabdoulline, R. R., and R. C. Wade. 2001. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol*. 306:1139–1155.
4. Baker N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041 2001.
5. Dolinsky T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker. 2004. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 32:W665–W667.
6. Dolinsky T.J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker. 2007. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 35:W522–W525.
7. Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem*. 98:1978–1988.
8. Wriggers, W., R. A. Milligan, and J. A. McCammon. 1999. Situs: a package for docking crystal structures into low resolution maps from electron microscopy. *J. Stuct. Biol*. 125:185–195.
9. Gabdoulline, R. R., and R. C. Wade. 1996. Effective charges for macromolecules in solvent. *J. Phys. Chem*. 100:3868–3878.
10. Elcock, A. H. 2006. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol*. 2:e98.
11. Gō, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng*. 12:183–210.

12. Frembgen-Kesner, T., and A. H. Elcock. 2009. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theor. Comput.* 5:242–256.

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

14. Rotne, J., and S. Prager. 1969. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys*. 50:4831–4837.

15. Yamakawa, H. 1970. Transport properties of polymer chains in dilute solution – hydrodynamic interaction. *J. Chem. Phys*. 53:436–443.

16. García de la Torre, J., and V. A. Bloomfield. 1977. Hydrodynamic properties of macromolecular complexes. I. Translation. *Biopolymers*. 16:1747–1763.

17. Carrasco, B., J. García de la Torre, and P. Zipper. 1999. Calculation of hydrodynamic properties of macromolecular bead models with overlapping spheres. *Eur. Biophys. J.* 28:510–515.

18. García de la Torre, J., M. L. Huertas, and B. Carrasco. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78:719–730.

19. García de la Torre, J. 2001. Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. *Biophys. Chem*. 93:159–170.

20. Allen, M.P., and D. J. Tildesley. 1989. Computer Simulation of Liquids. Oxford University Press Inc., New York.

21. Northrup, S. H., S. A. Allison, and J. A. McCammon. 1984. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys*. 80:1517–1524.

22. Gabdoulline, R. R., and R. C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.

23. Camacho, C. J., S. R. Kimura, C. DeLisi, and S. Vajda. 2000. Kinetics of desolvation-mediated protein-protein binding. *Biophys. J.* 78:1094–1105.

24. Gabdoulline, R. R., and R. C. Wade. 2001. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol.* 306:1139–1155.

25. Spaar, A., C. Dammer, R. R. Gabdoulline, R. C. Wade, and V. Helms. 2006. Diffusional encounter of barnase and barstar. *Biophys. J.* 90:1913–1924.

26. Clementi, C., Nymeyer, H., and J.N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? *J. Mol. Biol*. 298:937–953.

27. Koga, N., and S. Takada. 2001. Role of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol*. 313:17–180.