PROTEINS: Structure, Function, and Genetics 52:339–348 (2003)

Nonequilibrium, Multiple-Timescale Simulations of Ligand-Receptor Interactions in Structured Protein Systems

Ying Zhang,[1] Michael H. Peters,[1,*] and Yaohang Li[2]
[1]*Program in Biomedical Engineering, Florida State University, FAMU-FSU Joint College of Engineering, Tallahassee, Florida*
[2]*Department of Computer Science, Florida State University, Tallahassee, Florida*

ABSTRACT **Predicting the long-time, nonequilibrium dynamics of receptor-ligand interactions for structured proteins in a host fluid is a formidable task, but of great importance to predicting and analyzing cell-signaling processes and small molecule drug efficacies. Such processes take place on timescales on the order of milliseconds to seconds, so "brute-force" real-time, molecular or atomic simulations to determine absolute ligand-binding rates to receptor targets and over a statistical ensemble of systems are not currently feasible. In the current study, we implement on real protein systems a previously developed[3–5] hybrid molecular dynamics/Brownian dynamics algorithm, which takes advantage of the underlying, disparate timescales involved and overcomes the limitations of brute-force approaches. The algorithm is based on a multiple timescale analysis of the total system Hamiltonian, including all atomic and molecular structure information for the system: water, ligand, and receptor. In general, the method can account for the complex hydrodynamic, translational-orientational diffusion aspects of ligand-docking dynamics as well as predict the actual or absolute rates of ligand binding. To test some of the underlying features of the method, simulations were conducted here for an artificially constructed spherical protein "made" from the real protein insulin. Excellent comparisons of simulation calculations of the so-called grand particle friction tensor to analytical values were obtained for this system when protein charge effects were neglected. When protein charges were included, we found anomalous results caused by the alteration of the spatial, microscopic structure of water proximal to the protein surface. Protein charge effects were found to be highly significant and consistent with the recent hypothesis of Hoppert and Mayer (Am Sci 1999;87:518–525) for charged macromolecules in water, which involves the formation of a "water dense region" proximal to the charged protein surface followed by a "dilute water region." We further studied the algorithm on a D-peptide/HIV capside protein system and demonstrated the algorithms utility to study the nonequilibrium docking dynamics in this contemporary problem. In general, protein charge effects, which alter water structural properties in an anomalous fashion proximal to the protein surface, were found to be much more important than the so-called hydrodynamic interaction effects between ligand and receptor. The diminished role of hydrodynamic interactions in protein systems allows for a much simpler overall dynamic algorithm for the nonequilibrium protein-docking process. Further studies are now underway to critically examine this simpler overall algorithm in analyzing the nonequilibrium protein-docking problem. Proteins 2003;52:339–348.**

Key words: protein docking; nonequilibrium adsorption; molecular dynamics; Brownian dynamics

INTRODUCTION

The transport, diffusion, and binding of peptides or proteins (ligands) in solution with cell surface receptor proteins is a complex, nonequilibrium process of fundamental importance to understanding and predicting cell-signaling pathways and in the analysis of lead drug candidates for known protein targets. In general, the overall dynamics is dependent on the specific interatomic interactions between ligand and receptor, including topology, charge, van der Waals constants, and so forth, and the protein dynamic interactions with the host fluid, including hydrophobic and hydrogen bonding effects in water-based systems.

Because the local diffusion and binding process in protein-signaling or -docking systems is on the order of milliseconds to seconds, real-time, brute-force molecular dynamic simulations, with timesteps on the order of femtoseconds and total atomic numbers in the tens of thousands, are not currently feasible. For example, the recent literature on molecular dynamic studies in protein-water systems[1] shows that the maximum time currently accessible is on the order of microseconds for single proteins in water. This is a far cry short of that required to study nonequilibrium protein-protein interactions where it is necessary to study a statistical ensemble of thousands

Grant sponsor: National Science Foundation; Grant number: CTS-9633495.

*Correspondence to: Michael Peters, Program in Biomedical Engineering, Florida State University, FAMU-FSU Joint College of Engineering, Tallahassee, FL 32306. E-mail: peters@eng.fsu.edu

Received 30 August 2002; Accepted 31 December 2002

of interactions each on millisecond to second timescales. We also note that a significant number of studies on ligand and receptor equilibrium binding have also been presented, which have been termed the "protein docking problem."[2] Our interest here, however, is accounting for nonequilibrium effects, such as ligand orientational diffusion limitations and prediction of absolute ligand-binding rates; therefore, it can be viewed as complimentary to equilibrium approaches and analysis.

In three previous articles we developed a molecular, statistical mechanics approach to describe the long-time dynamics of macromolecules near external surfaces including the complete molecular detailed structure of the particle (ligand), surface (receptor), and solvent.[3–5] (Hereafter, they are referred to as I, II, and III, respectively.)

In general, our methods are based on a molecular derivation of the Fokker–Planck (FP) equation, which describes in a general fashion the dynamics of a structured Brownian particle immersed in a molecular solvent near a fixed surface. Specifically, the FP equation was derived in I for an arbitrary-shaped, molecularly structured Brownian particle near a molecularly structured external surface or "wall" by using the multiple timescales perturbation method. The derivation exploits the disparate timescales associated with the dynamics of a Brownian particle in a molecular fluid and leads to a fluctuation-dissipation relation for the grand particle friction tensor. In these types of relationships, also called Green–Kubo relations, the dynamic property is determined from an equilibrium time correlation function. Physically, one is studying the natural fluctuations in the instantaneous force exerted on the particle by the fluid under equilibrium conditions. Although such fluctuations take place over very small timescales, on the order of femtoseconds, as shown in II, it is possible to determine these effects in complex, arbitrarily structured systems with canonical, equilibrium molecular dynamics (MD) computational methods.

We have also shown in I and II that in the presence of an external surface, new terms appear in the derivation of the generalized FP equation as a result of fluid molecular mediated interactions between the particle and the surface (molecular counterpart of "hydrodynamic interactions"). It was also shown in II that MD calculations of the force autocorrelation function in simplified geometric systems lead to results in agreement with continuum hydrodynamics, except under conditions of particle-surface separation distances of a few molecules where the continuum assumption breaks down.

Now, in systems where the macromolecular particle "relaxation time" (proportional to the inverse particle friction tensor times the particle mass) is small compared to a characteristic timescale associated with its displacement by external forces, or possibly an initial state, the FP equation can be reduced to a much simpler Smoluchowski (Sm) equation in which the particle momenta no longer appears. The molecular based Sm equation, corresponding to the FP equation in I, was obtained in III by using formal asymptotic methods. In addition, because of the practical importance of complex systems, we also formally derived a Brownian dynamics algorithm for the numerical solution to the Sm equation. We showed how a combined molecular/Brownian dynamics method (MD/BD method) can be used to determine the long-time dynamics of rigid, molecularly structured macromolecules near structured surfaces in molecularly specified fluids.

Simply speaking, the key to the hybrid MD/BD numerical algorithm is that the host fluid relaxes to an equilibrium state in the potential field of the Brownian particle over very short times on the order of picoseconds or less. However, significant changes in the Brownian particle's position and orientation (say a few percent or less) take place on much longer timescales on the order of microseconds or longer. This allows us to set the particle diffusion properties using MD and time correlation analysis (the picosecond calculation) and then "leap ahead" in time moving the particle according to a simple six-dimensional (three positional and three orientational coordinates) BD timestep (the microsecond calculation). The entire process, MD followed by BD, is then repeated for the new configurational state.

The hybrid MD/BD method is currently theoretically restricted to rigid macromolecules and external surfaces, and internal vibrational modes are not allowed. Nonetheless, as treated more fully here, the above studies have notable significance in applications in biological engineering and physics involving the prediction of rates and molecular mechanisms of site-specific adsorption or attachment of macromolecules onto cells and other surfaces ("Receptor-Mediated Processes"). In these systems, the specific molecular structure and interactions of the macromolecule (ligand) and surface (receptor) are critical in determining the association kinetics. It is important to note that the efficiencies of receptor-mediated processes depend on both the dynamics and statics of the molecular interactions between particle, receptor, and host fluid. Equilibrium statistical thermodynamics only gives equilibrium-binding information and not dynamic (rate) information on the overall physical attachment process (diffusion, transport, and attachment), including the effects of the host fluid (hydrophobic and hydrophilic effects); these can only be determined through dynamic algorithms, such as the hybrid MD/BD algorithm given in III. The novelty of the MD/BD method is that it simulates receptor-mediated processes on timescales on the order of milliseconds to seconds and still retains the complete molecular details of the system. Such large timescale simulations are not currently possible by MD alone.

Our goals of the present work are to study the implementation of the hybrid MD/BD algorithm on several real protein systems. In the first system, we examine a spherically shaped protein artificially made from the protein insulin. The spherical shape allows us to compare with known analytical solutions for this system and thereby establish the efficacy of the method. The next system investigated is the interaction of a proposed D-peptide inhibitor drug with a segment of an HIV capsid protein. We examine some of the key features of this interaction and demonstrate how, in general, a complete "docking

dynamic simulation" of a peptide or small protein with a surface receptor protein can be made.

MATERIALS AND METHODS

Brief Review of the Hybrid MD/BD Method

In I, we derived the FP equation for a structured, rigid Brownian particle (B-particle) near a structured surface or wall beginning with the Liouville equation and the complete classical, atomistic Hamiltonian. The so-called multiple timescales perturbation method was then used to sort out the different underlying dynamics. The expansion is based on the usual Brownian particle assumption of a large mass ratio between the macromolecule and host fluid atom and results in a generalized FP equation applicable to $0[(m/M)^2]$, where m is the mass of a solvent particle and M is the mass of the macromolecule.

In III, we derived the Smoluchowski equation from the FP equation under the conditions where the particle relaxation time (defined as the B-particle mass divided by its characteristic friction coefficient) is small compared to a characteristic timescale associated with the particle displacement by an external force (or possibly and initial state), which is the typical case including protein-water systems (see III for more details).

Following formal asymptotic methods, in III we obtained the dimensionless Smoluchowski equation to $0(\epsilon^2)$ as

$$\frac{\partial n}{\partial t} = \epsilon[\partial/\partial \mathbf{R}(\partial/\partial\alpha \cdot \mathbf{\Lambda})]\begin{bmatrix} \mathbf{D}_T & \mathbf{D}_{TR} \\ \mathbf{D}_{RT} & \mathbf{D}_R \end{bmatrix}\begin{bmatrix} \partial n/\partial \mathbf{R} - n\mathbf{F}' \\ (\partial/\partial\alpha \cdot \mathbf{\Lambda})n - n\mathbf{T}' \end{bmatrix} + 0(\epsilon^2) \quad (1)$$

where $\mathbf{R}$ represents the position of the macromolecular center of mass, and the rotational operator, $(\partial/\partial\alpha \cdot \mathbf{\Lambda})$, is given by

$$(\partial/\partial\alpha \cdot \mathbf{\Lambda}) = \mathbf{e}_1\left[\cos\phi\,\frac{\partial}{\partial\theta} - \cot\theta\,\sin\phi\,\frac{\partial}{\partial\phi} + \csc\theta\,\sin\phi\,\frac{\partial}{\partial\psi}\right] + \mathbf{e}_2\left[\sin\phi\,\frac{\partial}{\partial\theta} + \cot\theta\,\cos\phi\,\frac{\partial}{\partial\phi} - \csc\theta\,\cos\phi\,\frac{\partial}{\partial\psi}\right] + \mathbf{e}_3\left[\frac{\partial}{\partial\phi}\right] \quad (2)$$

and ϕ, θ, and ψ are the Euler angles denoting the orientation of the macromolecule relative to a space-fixed or lab frame that has unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

The smallness parameter ϵ in Eq. (1) is given in terms of the macromolecular diffusion, $D_0 \equiv kT/\zeta_0$, as

$$\epsilon = \frac{D_0}{R_0(kT/M)^{1/2}} \equiv \frac{1}{N_{Pe}} \quad (3)$$

where k is Boltzmann's constant, T is temperature, ζ_0 is a characteristic macromolecular friction coefficient, and N_{Pe} is a type of Peclet number, which is the reciprocal ratio of the particle or macromolecular diffusion time scale (R_0^2/D_0) to the characteristic timescale $(t_0 = R_0/(kT/M)^{1/2})$ for particle translation or rotation due to the external force or torque field (or possibly an initial state).

The dimensionless particle diffusion tensors in Eq. (1) above are given in terms of the dimensionless particle friction tensors as (see III for details)

$$\mathbf{D}_T \equiv [\zeta_T - \zeta_{TR} \cdot \zeta_R^{-1} \cdot \zeta_{RT}]^{-1} \quad (4)$$

$$\mathbf{D}_{TR} \equiv [\zeta_{RT} - \zeta_R \cdot \zeta_{TR}^{-1} \cdot \zeta_T]^{-1} \quad (5)$$

$$\mathbf{D}_R \equiv [\zeta_R - \zeta_{RT} \cdot \zeta_T^{-1} \cdot \zeta_{TR}]^{-1} \quad (6)$$

$$\mathbf{D}_{RT} \equiv [\zeta_{TR} - \zeta_T \cdot \zeta_{RT}^{-1} \cdot \zeta_R]^{-1} \quad (7)$$

where the dimensionless translational, rotational, and coupled translational-rotational friction tensors are given by the time autocorrelation relationships (I and II)

$$\zeta_T \equiv \int_0^\infty [\langle \mathbf{F}_f(s)\mathbf{F}_f(0)\rangle_{eq} - \langle \mathbf{F}_f\rangle_{eq}^2]ds \quad (8)$$

$$\zeta_{TR} \equiv \int_0^\infty [\langle \mathbf{F}_f(s)\mathbf{T}_f(0)\rangle_{eq} - \langle \mathbf{F}_f\rangle_{eq}\langle \mathbf{T}_f\rangle_{eq}]ds \quad (9)$$

$$\zeta_{RT} \equiv \int_0^\infty [\langle \mathbf{T}_f(s)\mathbf{F}_f(0)\rangle_{eq} - \langle \mathbf{T}_f\rangle_{eq}\langle \mathbf{F}\rangle_{eq}]ds \quad (10)$$

$$\zeta_R \equiv \int_0^\infty [\langle \mathbf{T}_f(s)\mathbf{T}_f(0)\rangle_{eq} - \langle \mathbf{T}_f\rangle_{eq}^2]ds \quad (11)$$

where $\mathbf{F}_f(s)$ and $\mathbf{T}_f(s)$, respectively, are the force and torque acting on the Brownian particle by the fluid at time s. Furthermore, $\langle \mathbf{F}_f\rangle_{eq}$ and $\langle \mathbf{T}_f\rangle_{eq}$ are the equilibrium average force and torque, respectively, which as shown in I and II are not zero because of the presence of the wall.

The so-called generalized Einstein relationships, Eqs. (4)–(7), (in dimensionless terms above) are identical in form to those for an isolated Brownian particle.[6] The subtle difference lies in the defining equations for the friction tensors given by Eqs. (8)–(11). These expressions include the presence of the structured surface or wall through both the nonzero equilibrium average force and torque and the time-dependent force and torque acting on the particle by the fluid molecules, which are simultaneously in the presence of the potential field of the structured wall. In a continuum description, these wall effects lead to the so-called "hydrodynamic interactions" of the particle with the wall. Note that net force and torque acting on the Brownian particle, $\mathbf{F}'$ and $\mathbf{T}'$, include the direct particle-wall interaction forces and torques and the nonzero, equilibrium average fluid force and torque exerted on the particle (see III for more details and discussion).

Because of the complexities of molecularly structured systems and the numerous potential applications to complex, practical systems, solutions to the generalized Smoluchowski equation must be obtained numerically. Below we summarize a general Brownian dynamics algorithm based on the short-time solution of the Smoluchowski equation

that retains the specific molecular description of the system.

Brownian Dynamics Method

As shown previously by Ermak and McCammon,[7] a Brownian dynamics, numerical solution to the Smoluchowski equation can be obtained by deriving an analytical solution for its short-time behavior. In doing this, we consider that at time $t = 0$, the position and orientation of the Brownian particle are exactly known, that is,

$$n = \delta(\mathbf{R} - \mathbf{R}_0)\delta(\mathbf{\Omega} - \mathbf{\Omega}_0), \text{ at } t = 0 \tag{12}$$

where δ is the Dirac delta function and $\mathbf{\Omega}$ is short-hand notation for the set of Euler angles. Now, following Ermak and McCammon[7] for very small times t ($t > 0$), we can assume that over the time interval $(0, t)$ all spatial and orientational functions are approximately constant at their initial values, that is,

$$\mathbf{F}' \cong \mathbf{F}'^0 \tag{13}$$

$$\mathbf{T}' \cong \mathbf{T}'^0 \tag{14}$$

$$\mathbf{D}_J \cong \mathbf{D}_J^0, J = T, TR, RT, R \tag{15}$$

and so forth where the superscript (0) indicates the known initial values. In addition, with this approximation the rotational operator, $(\partial/\partial\alpha \cdot \mathbf{\Lambda})$, is given by

$$\begin{aligned}(\partial/\partial\alpha \cdot \mathbf{\Lambda})^0 &= \mathbf{e}_1\left[\cos\phi^0 \frac{\partial}{\partial\theta} - \cot\theta^0 \sin\phi^0 \frac{\partial}{\partial\phi} + \csc\theta^0 \sin\phi^0 \frac{\partial}{\partial\psi}\right] \\ &+ \mathbf{e}_2\left[\sin\phi^0 \frac{\partial}{\partial\theta} + \cot\theta^0 \cos\phi^0 \frac{\partial}{\partial\phi} - \csc\theta^0 \cos\phi^0 \frac{\partial}{\partial\psi}\right] \\ &+ \mathbf{e}_3\left[\frac{\partial}{\partial\phi}\right]\end{aligned} \tag{16}$$

Now we can readily show that the quantities in brackets in the above equation are the derivative operators for rotations about the space-fixed Cartesian frame (III, Appendix B; also see Goldstein[8])

$$(\partial/\partial\alpha \cdot \mathbf{\Lambda})^0 = \frac{\partial}{\partial\mathbf{\Phi}} \equiv \mathbf{e}_1 \frac{\partial}{\partial\phi_1} + \mathbf{e}_2 \frac{\partial}{\partial\phi_2} + \mathbf{e}_3 \frac{\partial}{\partial\phi_3} \tag{17}$$

Using Eq. (17), Eqs. (13)–(16), and neglecting spatial and orientational gradients in the diffusion tensors, the Smoluchowski equation, Eq. (1), can be rearranged to give, for small times,

$$\frac{\partial n}{\partial t} + \epsilon(\mathbf{D}_T^0 \cdot \mathbf{F}'^0 + \mathbf{D}_{TR}^0 \cdot \mathbf{T}'^0) \cdot \frac{\partial n}{\partial \mathbf{R}} + \epsilon(\mathbf{D}_{RT}^0 \cdot \mathbf{F}'^0 + \mathbf{D}_R^0 \cdot \mathbf{T}'^0) \cdot \frac{\partial n}{\partial \mathbf{\Phi}} = \epsilon[\partial/\partial\mathbf{R}\partial/\partial\mathbf{\Phi}][\mathbf{D}^0]\begin{bmatrix} \partial n/\partial\mathbf{R} \\ \partial n/\partial\mathbf{\Phi} \end{bmatrix} \tag{18}$$

where the 6×6 grand diffusion tensor, $\mathbf{D}^0$, is defined by

$$[\mathbf{D}^0] \equiv \begin{bmatrix} \mathbf{D}_T^0 & \mathbf{D}_{TR}^0 \\ \mathbf{D}_{RT}^0 & \mathbf{D}_R^0 \end{bmatrix} \tag{19}$$

TABLE I. Hybrid MD/BD Algorithm

1. Read standard PDB file for ligand and receptor.
2. Create "topology" file from PDB file (mass, charge, force constants).
3. Determine principal axes of ligand and initial Euler angles.
4. Hydrate ligand-receptor by using Monte Carlo and assign mass and charge to water (SPC model).
5. Perform equilibrium, canonical ensemble MD to get the grand friction tensor for ligand.
6. Numerically invert the grand friction tensor to obtain the grand diffusion tensor.
7. Perform BD to get new ligand position and orientation.
8. Go to 4.

Now, as shown in III, the above equation can be solved subject to the initial conditions, Eq. (12), and the result can put in the form of translational and rotational displacement equations as

$$R_i = R_i^0 + \epsilon t\left[\sum_j (D_{ij_T}^0 \mathbf{F}'^0_j + \mathbf{D}_{ij_{TR}}^0 T'^0_j)\right] + C_i(D_{ij}^0, t), \quad (1 \le i \le 3, 1 \le j \le 6) \tag{20}$$

$$\phi_i = \phi_i^0 + \epsilon t\left[\sum_j (\mathbf{D}_{ij_{RT}}^0 \mathbf{F}'^0_j + \mathbf{D}_{ij_R}^0 \mathbf{T}'^0_j)\right] + C_i(D_{ij}^0, t), \quad (4 \le i \le 6, 1 \le j \le 6) \tag{21}$$

where $C_i(D_{ij}^0, t)$ is a multivariate, Gaussian random number with zero mean and variance-covariance given by

$$\langle C_i(D_{ij}^0, t) C_j(D_{ij}^0, t)\rangle = 2D_{ij}^0 t\epsilon \tag{22}$$

Note that the indices on the components of the diffusion tensors in Eqs. (20) and (21) run according to Eq. (19).

Aside from the neglect of spatial and orientational gradients in the diffusion tensor, Eqs. (20) and (21) above are in agreement in form with that given previously by Dickinson et al.[9] for the combined rotational-translational dynamics of interacting Brownian particles. As we have noted in Brownian dynamics applications previously,[10] the torques and forces are usually expressed in terms of the Euler angles, so that conversions from rotations about the space-fixed Cartesian frame are needed (see III, Appendix B)

$$\theta - \theta^0 = \cos\phi^0 \Delta\phi_x + \sin\phi^0 \Delta\phi_y \tag{23}$$

$$\phi - \phi^0 = -\cot\theta^0 \sin\phi^0 \Delta\phi_x + \cot\theta^0 \cos\phi^0 \Delta\phi_y + \Delta\phi_z \tag{24}$$

$$\psi - \psi^0 = \csc\theta^0 \sin\phi^0 \Delta\phi_x - \csc\theta^0 \cos\phi^0 \Delta\phi_y \tag{25}$$

Implementation and Computational Strategies for Protein-Water Systems

The general computational MD/BD algorithm is shown in Table I. The computational scheme begins by reading a standard PDB file from the protein data bank for both ligand and receptor.[11] This file is then converted to a "topology" file that includes computationally critical information on atomic mass, residue charge, and Lennard–

Jones (LJ) interaction force constants with water oxygen. For the LJ force constants, we use those values given in GROMACS.[12] Next, the ligand and receptor must be hydrated by using a Monte Carlo method. Here we have chosen the previously developed hydration code SOLVATE.[13] Solvate accepts pdb-files of the protein as input, and the output is a pdb-file of the solute plus water molecules. For the molecular model of water, we use the so-called modified SPC model[14] with long-range electrostatic interatomic interactions accounted for by a modified Poisson–Boltzmann reaction field method.[15] The center of mass and body-fixed axes along the principal axes of inertia for the ligand are initially computed by standard formulas. This sets the body-fixed coordinates and initial Euler angles, the latter of which give the orientation of the body relative to the space-fixed frame. MD is then used to determine the particle grand friction tensor as discussed in I, II, and III. The grand friction tensor is numerically inverted to obtain the grand diffusion tensor. The grand diffusion tensor is then used to perform the BD move on a timestep of around 10^{-5} seconds. The macromolecule position and orientation change by only a couple of percent or less over this time period. The new atomic positions are updated based on the BD move and the entire process (i.e., MD followed by BD) is repeated.

For all results shown below, the MD timesteps were taken as 0.5 fs, and the autocorrelations were calculated over 3000 timesteps. Twenty ensembles were used to obtain the final average values and statistical uncertainties (see II for more details) resulting in 60,000 total MD timesteps for a particular protein system configuration. In all cases, the temperature of the system was taken as 298K. As before (II), the side dimension of the cubical MD box was always taken to be twice the particle (protein or protein-protein complex) diameter. All results shown here were obtained on a DEC Alpha DS10 6/466 with 256 of DRAM and 8 processors.

It is interesting to note here that the autocorrelation calculations, which form the basis of the more computationally expensive MD part of the algorithm, are particularly amenable to multiprocessor systems and the use of message-passing interface (MPI) programs. We have repeated many of the calculations shown here on such systems, and a general review of these large multiprocessing systems can be found in Ref. 16. In MPI application, each processor represents one member of the ensemble allowing hundreds of ensembles to be included. The 3000 MD timesteps to construct a single autocorrelation are done on each processor with a very minimal execution time of a few hours or less depending on each processor speed. The results from each processor are summed by using standard MPI instructions to obtain final average values and statistical uncertainties.

RESULTS AND DISCUSSION

Results for Single, Isolated Proteins in Water

To quantitatively test the methods and results, a spherically shaped protein was constructed, simulated, and compared to well-known analytical results for a sphere in water. To obtain a spherical protein, the protein insulin was selected from the protein data bank. It was then cut to be spherical by removing the atoms whose sum of their Cartesian coordinates squared was greater than the square of an assigned radius, which was arbitrarily chosen to be 12 Å. Because of the compact structure of insulin, this resulted in an overall spherically shaped protein with a "rough" surface as shown in Figure 1. For these initial studies, protein atomic hydrogens were neglected for the sake of simplicity.

Figure 2 shows the xx-component of the translational force autocorrelation function obtained from MD for the spherical insulin particle neglecting the atomic charges of the protein itself. Protein charge effects are considered separately below. Standard deviations in the average values are also shown in Figure 2, which can be used to determine confidence intervals and associated errors in computational results (also see II).

A confidence interval in the autocorrelation values, CI, is obtained from the Tchebycheff inequality as[18]

$$CI = 1 - \frac{\sigma^2}{n_r \epsilon^2} \tag{26}$$

where ϵ is the error in the autocorrelation, σ^2 is the variance of the autocorrelation, and n_r is the number of repeats. Because the friction tensor is the integral of the force autocorrelation, a good estimate of the error in a friction tensor element can be obtained by taking specific values at the midpoint of the decay. For example, in Figure 2, the average variance is approximately 100 and at the midpoint of the decay the force autocorrelation value is around 15. Thus, with $n_r = 20$, a 90% confidence interval leads to an error of around 50%. Note that the error is proportional to the reciprocal square root of the ensemble number, that is,

$$\epsilon \sim \frac{1}{n_r^{1/2}} \tag{27}$$

For example, in the above system, increasing the ensemble size to 40 would result in an error of around 30% for a 90% confidence. Our results show that at least 20 ensembles are minimally necessary in these systems, and more (perhaps double) would be desirable for more accurate results. Below, we also indirectly demonstrate a sufficient number of repeats by comparing our results to analytical solutions. As noted above, the autocorrelation calculations are particularly amenable to MPI routines where each processor represents one member of the ensemble allowing for hundreds of members leading to very accurate results.

In comparisons of MD results to analytical, continuum expressions, it has been previously shown that in "small" particle systems an "effective" sphere radius should be used in comparison tests based on the fluid volume excluded by the sphere.[19] For spherical insulin, the effective particle radius a_{eff} is, therefore,

$$a_{\text{eff}} \cong a + \frac{\sigma_{oo}}{2} \tag{28}$$

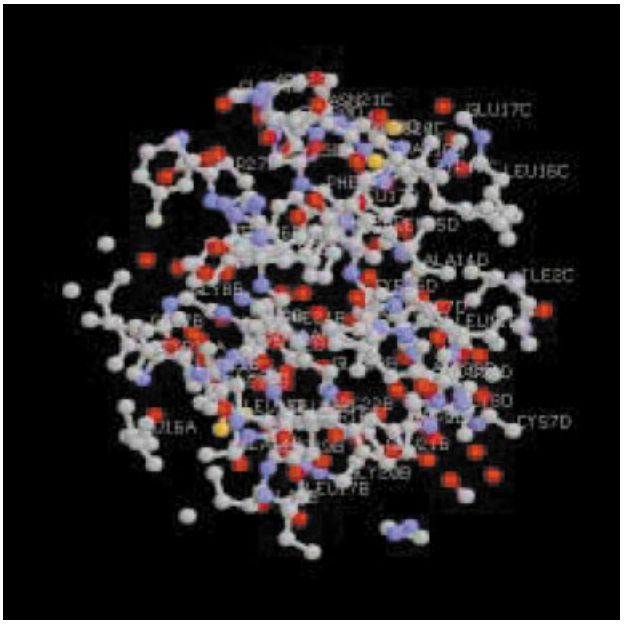

Fig. 1. A "spherical" protein obtained from the real protein insulin. Illustration is generated by RASMOL.[17]

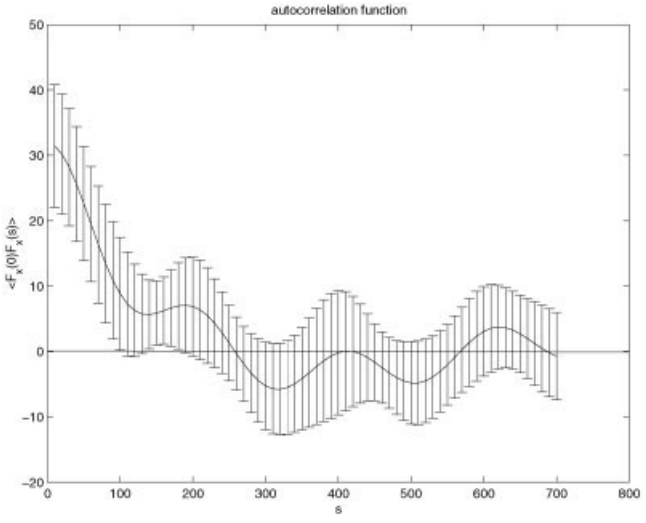

Fig. 2. The xx-component of the force autocorrelation function for spherical insulin neglecting protein charge effects. All quantities are dimensionless; a value of s = 100 corresponds to a real time of 0.05 ps.

where a is the spherical radius (12 Å) and the water molecular diameter, σ_{oo}, is 3.16 Å.

The analytical translational and rotational friction tensors are well known as[20]

$$\zeta_T = 6\pi\mu a_{\text{eff}}(1 - \xi) \tag{29}$$

$$\zeta_R = 8\pi\mu a_{\text{eff}}^3(1 - 3\xi) \tag{30}$$

where ζ_T is the translational friction coefficient, ζ_R is the rotational friction coefficient, μ is viscosity (1.0 centipoise at 20°C); ξ is a slip coefficient[19] with $\xi = 0$ being the no-slip or "stick" condition and $\xi = \frac{1}{3}$ is the full slip condition.

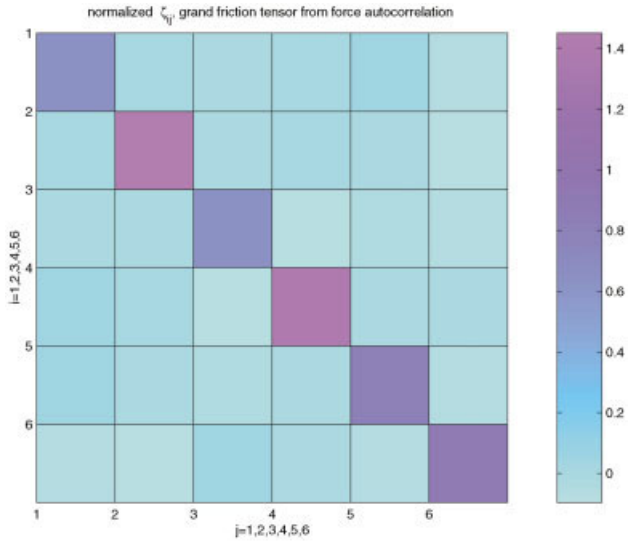

Fig. 3. The normalized grand friction tensor for spherical insulin: pure translational *i*, *j*ϵ1, 2, 3; pure rotational *i*, *j*ϵ4, 5, 6; translational and rotational *i*ϵ1, 2, 3, *j*ϵ4, 5, 6, or *i*ϵ4, 5, 6, or *i*ϵ4, 5, 6, *j*ϵ1, 2, 3.

For a small, aggregate sphere, previous MD results[19] showed that the slip coefficient is between the limits of full slip and stick boundary conditions. For the spherical insulin system, a slip coefficient of 0.27 was found to give the best comparison to MD-generated values of the friction tensor as shown below.

The MD generated numerical values of each of the elements of the 6×6 symmetric[21] grand friction tensor for spherical insulin, normalized as $(\zeta^{T}/\zeta^{T}_{\text{anal}})$, $(\zeta^{R}/\zeta^{R}_{\text{anal}})$, $(\zeta^{TR}/\zeta^{R}_{\text{anal}})$, are tabulated below and shown graphically in Figure 3. Note that the slip coefficient of 0.27 used for analytical comparisons was closer to that for slip conditions because of the highly irregular nature of the protein surface (Fig. 1).

0.64	0.0014	−0.0002	0.0068	0.052	−0.061
	1.45	−0.0033	0.012	−0.021	−0.076
		0.64	−0.096	−0.042	0.063
			1.38	−0.0001	−0.022
				0.80	−0.049
					0.89

The diagonal elements of the grand friction tensor are seen to be within experimental error of the analytical results for an isotropic, spherical particle. Note that the off-diagonal elements of the grand friction tensor show even better agreement with the analytical result in that they are predicted to be within a few percent of zero.[20] This prediction of the entire behavior of the grand particle friction tensor is an important feature when such methods are applied to disparately shaped proteins and peptides where the coupled translational-rotational dynamics can, in general, significantly affect particle dynamics.[21]

Figure 4 shows the water radial distribution function for the uncharged spherical insulin system. The biggest peak at 5 (dimensionless r) is the first coordination layer of the water molecules outside the protein; the next peak is the

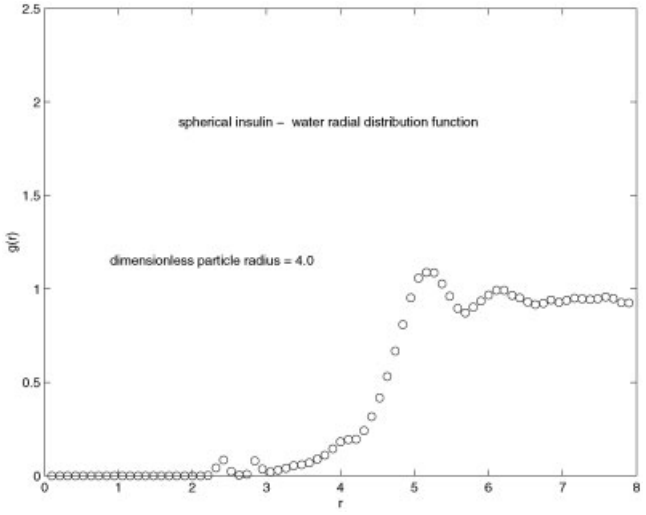

Fig. 4. Water radial distribution function for the spherical insulin system. All quantities are dimensionless.

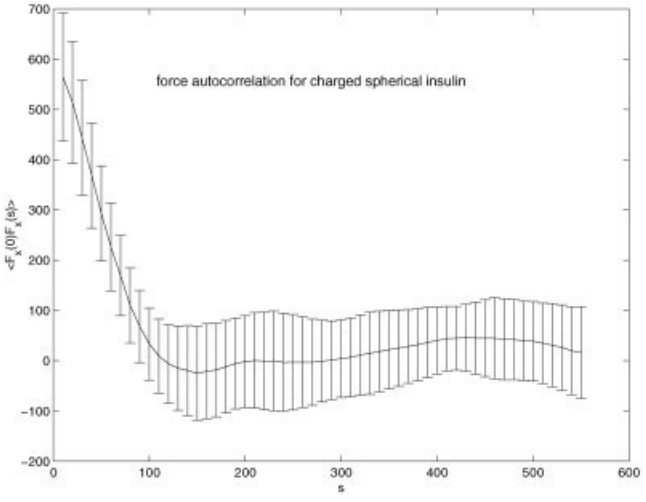

Fig. 5. xx-Force autocorrelation function for the charged spherical insulin system. All quantities are dimensionless.

second coordination layer, and so forth. The curve ends up with a platform close to the water bulk density. The small peaks between 2 and 3 are those water molecules that become trapped or partially trapped inside of the protein. The results for the uncharged particle are very similar to our previous results for atomic BCC or FCC packing, spherical particles (II). The coordination peaks in Figure 4 are not quite as sharp and steep as in II because of the highly irregular nature of the spherical protein surface, but otherwise they are very similar.

Protein Charge Effects

We repeated all of the above calculations for spherical insulin, including the protein atomic charges. Figure 5 shows the x-x force autocorrelation function, and Figure 6 displays the radial distribution function. The numerical values of the 6×6 grand friction tensor were obtained as

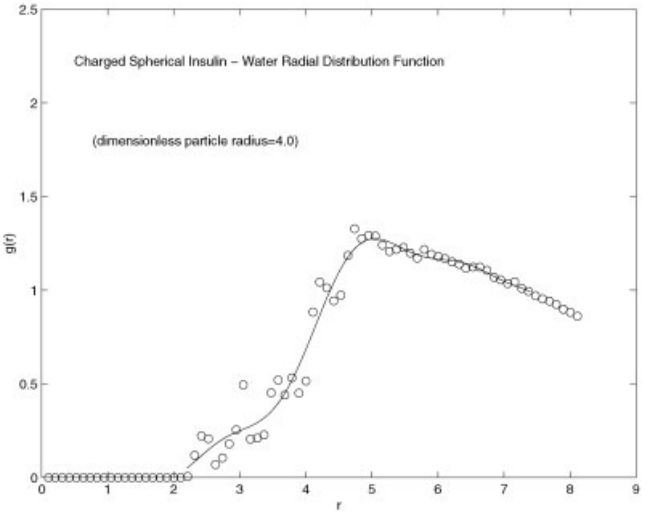

Fig. 6. Water radial distribution function for the charged spherical insulin system. All quantities are dimensionless.

$$\begin{matrix} 9.80 & -1.47 & -3.86 & 0.13 & -0.05 & -0.10 \\ & 6.93 & 4.11 & 0.22 & 0.23 & 0.02 \\ & & 25.13 & 1.78 & 0.29 & 0.90 \\ & & & 10.86 & -0.60 & 0.23 \\ & & & & 6.36 & -1.45 \\ & & & & & 11.57 \end{matrix}$$

where we used the identical normalization as in the uncharged case. It can be seen from these values that the diagonal friction tensor elements have increased by severalfold over the uncharged values given previously. In addition, both the autocorrelation function (Fig. 5) and radial distribution function (Fig. 6) have changed their behavior considerably. In particular, the second coordination peak in the uncharged protein case (at a dimensionless radius of about 6 in Fig. 4) has disappeared. There is now only a gradual decline in water density from its peak value at the first coordination layer. The water density gradually crosses below the $g(r) = 1.0$ line at about $r = 7$ in Figure 6. This observed behavior in the radial distribution function, i.e., a "water dense" region near the surface (as opposed to two or more coordination peaks), followed by a "less dense" region is consistent with the recent hypothesis of Hoppert and Mayer[22] for charged macromolecules in water. Those authors suggested that charged macromolecules may have "more dense" and "less dense" water regions about their surface owing to the water dipole interactions with the surface charges. Note, by comparison of Figures 4 and 6, that the effective particle radius is similar in both the uncharged and charged cases, and the differences must be attributed to the behavior of the water-protein force autocorrelation, which in the case of charged proteins leads to a larger zero-time magnitude ("static autocorrelation value") and a much sharper decay or shorter lived force correlations. Macroscopically, such changes can be attributed to local fluid viscosity increases near the charged protein surface through the fluid stress autocorrelation function,[20] albeit we have not performed

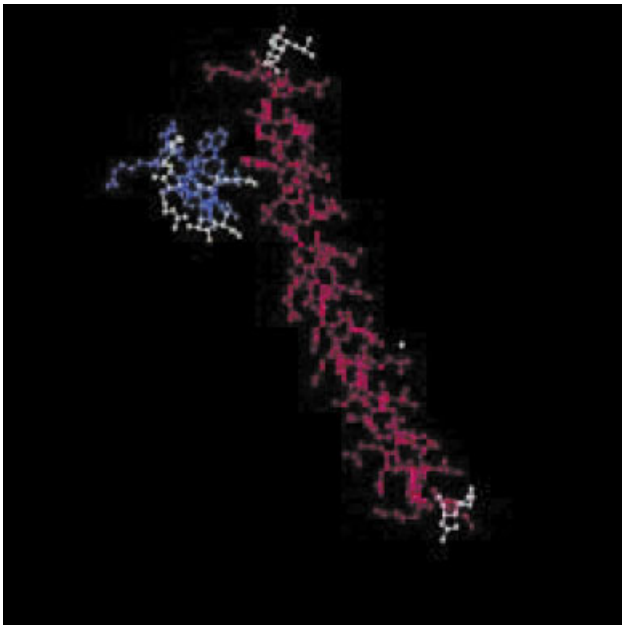

Fig. 7. Molecular structure of the D-peptide/gp41 complex. Note that the D-peptide has been moved several angstroms from the gp41 surface in this picture.

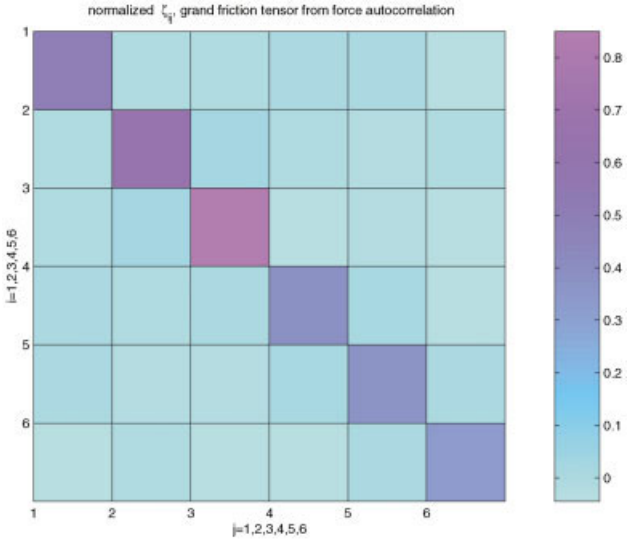

Fig. 8. The grand friction tensor of uncharged D-peptide: pure translational $i, j\epsilon 1, 2, 3$; pure rotational $i, j\epsilon 4, 5, 6$; translational and rotational $i\epsilon 1, 2, 3, j\epsilon 4, 5, 6$, or $i\epsilon 4, 5, 6, j\epsilon 1, 2, 3$.

those specific fluid stress autocorrelation calculations here. However, significant changes in the water self-diffusion coefficient in charged protein solutions from the "free" water value have also been experimentally reported by Lamanna et al.[23] using NMR, and these changes were also attributed by those authors to the alterations in water microstructure near the charged protein surface. Computationally, we have found that protein charge effects can be highly significant in determining the absolute magnitudes of translation and orientational friction or mobilities of proteins in water. More detailed studies are underway to further examine these induced microstructural changes of water, including time-averaged dipole orientations, protein atomic hydrogen effects, water stress autocorrelation function analysis, and full characterization of the less dense region, the latter of which requires a larger MD box size than the one shown here. Although not shown here for the sake of brevity, increasing the size of the MD box to capture more of the less dense water region showed no effect on the friction or mobility calculations given above. It is still quite interesting that the protein charges alter the water molecular distribution at relatively large distances from the protein surface, which could have other applicational significance such as in water microstructure in cell cytoplasma.

Results for D-Peptide

Recently,[24] a particular form of a D-peptide, based on D-amino acid isometric forms, was proposed as an inhibitor for the entry of HIV into host cells. The D-peptide binds to a segment of the HIV capsid protein called gp41. The complete molecular structure of this D-peptide/gp41 complex has been determined[24] (see Fig. 7), and it represents a contemporary, real protein system to study the underlying features of the MD-BD code.

Before analyzing the dynamics of the gp41/D-peptide complex, however, the D-peptide alone was simulated. For analytical comparison purposes, we can obtain the effective peptide diameter (d_{eff}) from the peptide volume (pepvol), where peptide volume = box volume − water volume. The box length in this system is 36 Å, the number of water molecules is 1403, and the water bulk density is $3.337 * 10^{28}$ (number/m^3). From pepvol $= \frac{1}{6} * \text{pi} * d_{\text{eff}}^3$ we obtain the effective spherical radius a_{eff} as 10 Å.

Figure 8 shows the normalized grand friction tensor for the uncharged D-peptide using the effective spherical radius for the analytic value. It can be seen that the D-peptide is very much spherelike in its friction dynamics owing to its compact, and near sphere-like shape, although some off-diagonal elements exist.

Figure 9 shows the water radial distribution function for the uncharged D-peptide. In comparison with the water distribution function of the spherical insulin, we can see that the coordination layers have disappeared because of the slightly nonspherical shape and highly diffuse topology of the peptide surface.

Although not shown here for the sake of brevity, we found similar anomalous changes for the charged D-peptide as observed above for the charged spherical insulin system. The diagonal elements of the grand friction tensor increased by several hundred percent over the uncharged D-peptide case, and there was again a major alteration in the radial distribution function behavior with a dense and dilute water regimen just outside of the peptide surface.

gp41/D-Peptide Complex Simulations and Implications for Brownian Dynamics

We conducted several MD simulations to obtain the grand friction tensor for the gp41/D-peptide complex sys-

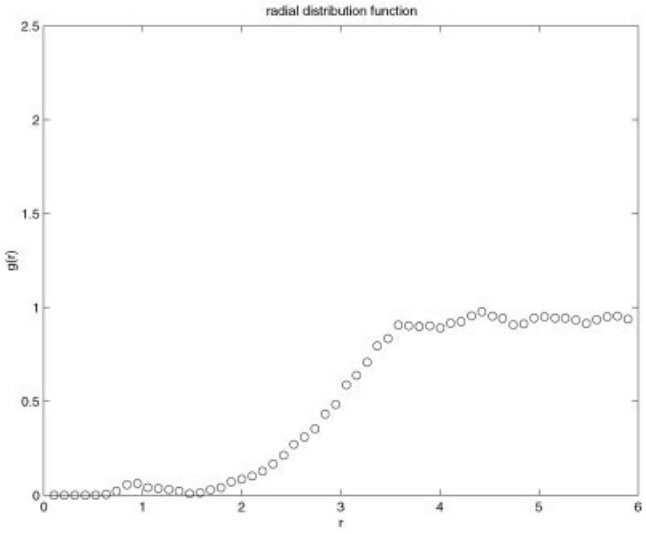

Fig. 9. Water radial distribution function in the D-peptide system.

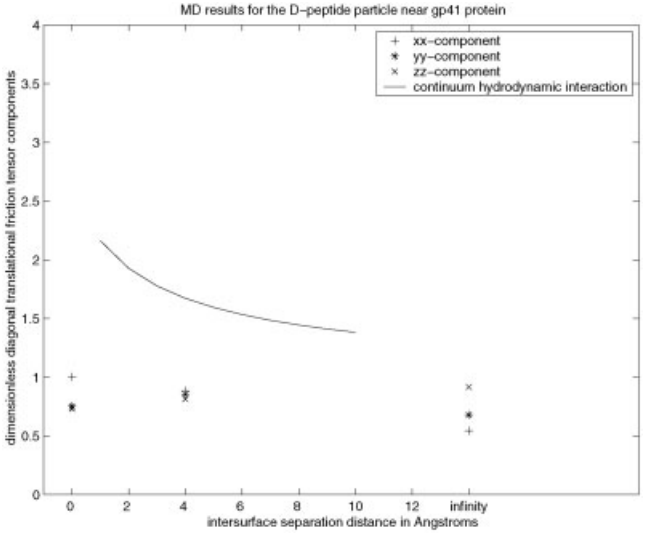

Fig. 10. Translational components of the normalized grand friction tensor for the D-peptide particle near the gp41 surface protein.

tem (Fig. 7) described above. Of most interest is the so-called hydrodynamic interaction effect whereby the grand particle (D-peptide) friction tensor increases because of the presence of another surface (gp41). Figure 10 shows some of the results of the diagonal components of the grand friction tensor for the D-peptide particle near the gp41 surface predicted by MD and autocorrelation analysis as a function of intersurface separation distances. The intersurface separation value of zero shown in Figure 10 corresponds to the original bound structure of the D-peptide to gp41, and the intersurface separation of infinity corresponds to the free D-peptide values given previously. The intermediate intersurface separation was obtained by artificially moving the entire D-peptide outward from its bound state with gp41.

The results shown in Figure 10 are consistent with our earlier findings for very small particles with diameters on the order of nanometers (II). Quantitatively from continuum hydrodynamic analysis of a spherical particle near a plane wall,[21] the translational or rotational friction coefficients would increase significantly as the particle nears the surface within a distance roughly corresponding to its diameter, as shown in Figure 10. In these very small particle systems, however, the hydrodynamic interaction effects expected from continuum fluid mechanics are diminished because of a breakdown of the continuum hypothesis at close intersurface separations (II). Although not shown here for the sake of brevity, similar small hydrodynamic interaction effects were observed for the charged protein system.

CONCLUSIONS

In general, our findings suggest that the large alterations in the frictional resistance or mobility of nanosized protein particles due to its anomalous charge interactions with water significantly outweigh any alterations associated with the so-called hydrodynamic interaction effects. This observation has important implications to the MD-BD code for charged protein interactions as explained more fully below.

For rigid macromolecular ligands in the absence of any other surface receptor site (i.e., in an infinite medium), the MD calculation has to be performed only once because the grand diffusion tensor is now invariant in the body frame of the ligand. This, of course, is not generally true when an external surface is present due to hydrodynamic interaction effects. However, our results here show that it is much more important to account for protein charge effects in the MD calculations of the grand friction tensor rather than hydrodynamic interaction effects due to the presence of the protein receptor. Thus, as a good approximation, the grand friction tensor calculated for the isolated charged protein or peptide ligand can be used in the BD simulation of the docking dynamics and, thus, MD has to be performed only once at the start of the BD simulation. The grand friction tensor can be numerically inverted with ease to obtain the grand diffusion tensor, and the transformations of the grand diffusion tensor from the invariant body-fixed frame to the laboratory frame are accomplished by standard transformation matrices in terms of the Euler angles (III). For the nonequilibrium protein-docking problem, the simplifications noted above are highly significant because determining the association kinetic constants for any particular system requires the study of a statistical ensemble of protein-protein interactions. We have shown how each member of this ensemble can be simulated with a simple 12-dimensional, long-time step Brownian dynamics algorithm [6 dimensions for each protein: 3 translational and 3 rotational; see, e.g., Ref. (10) for more details on the BD simulation procedure]. This simpler algorithm is currently being tested on some small protein, receptor-ligand systems, and it has practical significance in the development of novel data-mining methods for lead ligand identification against known protein targets or receptors.

We also note that it is possible to extend the underlying theories used here for rigid proteins to flexible proteins

using small vibration theory approaches,[25] because intramacromolecular motions in proteins take place on timescales greater than the picosecond or less timescale necessary to establish a Fokker–Planck description (see I and II for more details); this promising theoretical work is currently underway.

REFERENCES

1. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744; Duan Y, Kollman PA. State of the art in studying protein folding and protein structure prediction using molecular dynamics methods. J Mol Graph Model 2001;19:146–149.
2. For some recent equilibrium binding studies on the protein docking problem, see, for example, Gabb HA, Jackson RM, Sternberg MJE. Modeling protein docking using shape complementarity, electrostatics, and biochemical information. J Mol Biol 1997; 272:106–120; Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 1998;19:1639–1662; and Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein-protein docking. Proteins 2000;40:525–537.
3. Peters MH. Fokker-Planck equation and the grand molecular friction tensor for coupled rotational and translational motions of structured Brownian particles near structured surfaces. J Chem Phys 1999;110:528–538.
4. Peters MH. Fokker-Planck equation, molecular friction, and molecular dynamics for Brownian particle transport near external solid surfaces. J Stat Phys 1999;94:557–573.
5. Peters MH. The smoluchowski equation for structured macromolecules near structured surfaces. J Chem Phys 2000;112:5488–5498.
6. Condiff DW, Dahler JS. Brownian motion of polyatomic molecules: the coupling of rotational and translational motions. J Chem Phys 1966;44:3988–4005.
7. Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. J Chem Phys 1978;69:1352–1360.
8. Goldstein H. Classical mechanics. Boston, MA: Addison-Wesley; 1980.
9. Dickinson E, Allison SA, McCammon JA. Brownian dynamics with rotational-translational coupling. J Chem Soc Faraday Trans 2 1985;81:591–601.
10. Peters MH, Ying R. Rotational and translational dynamics of flexible macromolecules. J Chem Phys 1993;98:6492–6503.
11. Berman HM, Westbrook J, Feng Z, Gililand G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
12. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message passing parallel molecular dynamics implementation. Comp Phys Comm 1995;91:43–56; Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Mod 2001;7:306–317.
13. See Grubmuller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996;271:997–1000; and the references cited therein.
14. Toukau K, Rahman A. Molecular dynamics study of atomic motions in water. Phys Rev B 1985;31:2643–2648.
15. Tironi IG, Sperb R, Smith PE, van Gunsteren WW. A generalized reaction field method for molecular dynamics simulations. J Chem Phys 1995;102:5451–5459.
16. Gropp W, Lusk E, Skjellum A. Using MPI. Cambridge, MA: MIT Press; 1999.
17. Sayle R, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci 1995;20:374.
18. Papoulis A. Random variables and stochastic processes, New York: McGraw-Hill; 1965.
19. Alder BJ, Alley WE. Generalized hydrodynamics. Phys Today 1984;January:56–63.
20. Schmitz R, Felderhof BU. Creeping flow about a sphere. Physica 1978;92A:423–437.
21. In computing the off-diagonal elements of the grand friction tensor for an isolated protein or peptide, we have used the well-known, general symmetric properties of this tensor for an arbitrarily shaped, isolated particle. Specifically, in computing the autocorrelation values for the spherical insulin and D-peptide systems, the (i, j) components and (j, i) components ($j \neq i$) were averaged, leading to the exact symmetrical tensor values shown. See Brenner H. J Coll Interface Sci 1967;23:407; Happel J, Nijhof; Brenner H. Low Reynolds number hydrodynamics. Dordrecht, Netherlands 1983.
22. Hoppert M, Mayer F. Prokaryotes. Am Sci 1999;87:518–525.
23. Lamanna R, Delmelle M, Cannistraro S. Solvent Stokes-Einstein violation in aqueous protein solutions. Phys Rev E 1994;49:5878–5880.
24. Eckert D, Nalashkevich V, Hong L, Carr P, Kim P. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target gp41 coiled coil pocket. Cell 1999;99:103–115.
25. Wilson E Bright Jr, Decius JC, Cross PC. Molecular vibrations, New York: McGraw-Hill; 1955.