JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Calculating Position-Dependent Diffusivity in Biased Molecular Dynamics Simulations

Jeffrey Comer,[†,‡] Christophe Chipot,*[,§,∥] and Fernando D. González-Nilo[†,‡,⊥]

[†]Fraunhofer Chile Research, Mariano Sánchez Fontecilla 310 piso 14, Las Condes, Santiago, Chile

[‡]Universidad Andres Bello, Center for Bioinformatics and Integrative Biology, Facultad de Ciencias Biológicas, Av. República 239, Santiago, Chile

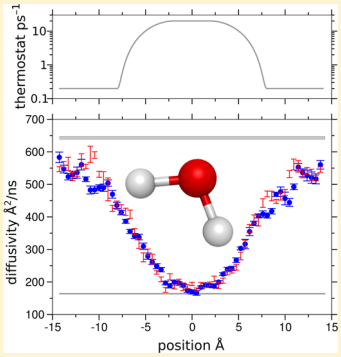
[§]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, 405 North Mathews, Urbana, Illinois 61801, United States

[∥]Équipe de dynamique des assemblages membranaires, UMR 7565, Université de Lorraine, BP 239, 54506 Vandœuvre-lès-Nancy cedex, France

[⊥]Centro Interdisciplinario de Neurociencia de Valparaíso, Universidad de Valparaíso, Valparaíso, Chile

Supporting Information

ABSTRACT: Calculating transition rates and other kinetic quantities from molecular simulations requires knowledge not only of the free energy along the relevant coordinate but also the diffusivity as a function of that coordinate. A variety of methods are currently used to map the free-energy landscape in molecular simulations; however, simultaneous calculation of position-dependent diffusivity is complicated by biasing forces applied with many of these methods. Here, we describe a method to calculate position-dependent diffusivities in simulations including known time-dependent biasing forces, which relies on a previously proposed Bayesian inference scheme. We first apply the method to an explicitly diffusive model, and then to an equilibrium molecular dynamics simulation of liquid water including a position-dependent thermostat, comparing the results to those of an established method. Finally, we test the method on a system of liquid water, where oscillations of the free energy along the coordinate of interest preclude sufficient sampling in an equilibrium simulation. The adaptive biasing force method permits roughly uniform sampling along this coordinate, while the method presented here gives a consistent result for the position-dependent diffusivity, even in a short simulation where the adaptive biasing force is only partially converged.



INTRODUCTION

Free-energy calculations in molecular simulations have become a workhorse of modelers studying biomolecular and other nanoscale systems. However, free energies provide only limited information about system dynamics, and additional information is required for the determination of kinetic quantities such as transition rates, permeation coefficients, particle currents, and mean first-passage times. As originally described by Kramers,[1] such quantities can be calculated by employing models of diffusive motion.

In the simplest diffusive model, one ignores inertia and memory effects and regards the diffusive motion as the random walk of a particle under a position-dependent potential.[2] Under the assumptions of this model, the dynamics of the coordinate of interest, $q(t)$, are determined by two functions:[3,4] $W(q)$, the potential of mean force along this coordinate (PMF; i.e., the free energy as a function of the coordinate with all other microscopic degrees of freedom taking on their thermodynamic averages), and $D(q)$, the diffusivity along this coordinate. In general, $D(q)$ is position-dependent (depends on q) and is likely to vary substantially in heterogeneous systems, wherein $W(q)$ is likely to vary substantially as well.

A variety of methods are commonly used for computing $W(q)$, which depend on the system being studied, as well as personal preferences. Some of these methods employ biasing forces to overcome the limits of Boltzmann sampling in which energy barriers and transition states may be poorly sampled or not sampled at all.[5,6] Popular methods include a combination of "umbrella sampling" and the weighted histogram analysis method,[7,8] nonequilibrium work methods[9,10] and the related forward and reverse (FR) pulling method,[4] adaptive force biasing (ABF),[11,12] and metadynamics.[13] Aside from umbrella sampling, all these methods involve the application of time-dependent biasing potentials or forces.

Like $W(q)$, calculation of $D(q)$ requires good sampling over the range of q of interest. For this reason, and for the sake of efficiency, it makes sense to combine the calculation of $D(q)$ and $W(q)$. Furthermore, as noted by Hummer, developing a correct diffusive model requires that $W(q)$ and $D(q)$ be determined self-consistently. For example, the appropriate $D(q)$ can depend on the spatial resolution of $W(q)$: a rough PMF at a

Received: October 7, 2012
Published: January 7, 2013

fine length scale can be modeled by a smoother PMF and a reduced diffusivity at a coarser length scale.[14] However, combining diffusivity calculations with PMF calculations is made difficult by the presence of the biasing forces, which assist in the estimation of the free energy change.

Depending on the PMF calculation method employed, diffusivities have sometimes been calculated from the same trajectories as were used from the PMF calculation and, other times, have been calculated in separate simulations. The most well-known relation for calculating diffusivity is $D = \lim_{t\to\infty} \langle [q(t) - q(0)]^2 \rangle/(2t)$; however, this relation is not even a good approximation in regions where $W(q)$ varies by more than k_BT.[15] Computing the diffusivity by the velocity autocorrelation function yields similar problems.[15] Among PMF calculation methods, the FR method is distinguished by the fact that a formula for $D(q)$ arises more or less naturally from the formalism.[4] For simulations in which the coordinate of interest is constrained, $D(q)$ may be calculated from the force autocorrelation function.[16] Due to the fact that harmonic potentials are often used in umbrella sampling, a method based on the generalized Langevin equation for a harmonic oscillator (first described by Woolf and Roux[17] and later reformulated by Hummer[3]) has been used to obtain the diffusivity.[18] However, the sum of the biasing harmonic potential and the intrinsic PMF is not exactly harmonic, and a stronger harmonic potential than that used in umbrella sampling may be needed to accurately calculate the diffusivity. Thus, separate simulations may be necessary[19] to employ methods based on harmonic oscillators. In past studies employing ABF to obtain PMFs, separate simulations were performed to estimate position-dependent diffusivities.[20,21]

An approach for consistent calculation of $W(q)$ and $D(q)$ using Bayesian inference has been developed by Hummer;[3] however, it assumes detailed balance, which is violated by time-dependent biasing forces. Türkcan et al.[22] recently described another method applying Bayesian inference for consistently determining $D(q)$ and the position-dependent force, $f(q)$. This force may or may not be the gradient of the intrinsic potential, $f = -\nabla W$ (we use ∇W for dW/dq, despite the one-dimensional setting). To consistently and simultaneously calculate $W(q)$ and $D(q)$ in metadynamics or ABF simulations (or possibly other nonequilibrium simulations), a method capable of calculating the $D(q)$ for a coordinate under the influence of known time-dependent forces, $f_{bias}(t)$, is needed.

Here, we describe a modification of the method of Türkcan et al., which permits the consistent calculation of $W(q)$ and $D(q)$ while also handling known time-dependent biasing forces and, furthermore, maintains consistency with information about $W(q)$ (if any) already determined by the biasing method. Moreover, we utilize the form of the Brownian integrator[23] appropriate for systems with spatially dependent diffusion to calculate the Bayesian posterior probability. This form includes a term involving $\nabla D(q)$, neglected in ref 22, which we show improves the accuracy of the predicted system force, $f_{sys}(q) = -\nabla W(q)$.

THEORY

In their seminal paper on Brownian dynamics,[23] Ermak and McCammon derive a discrete integrator describing displacements of a Brownian particle under the influence of an arbitrary force and an arbitrary diffusivity tensor (eq 15 of Ermak and McCammon). For simplicity, we assume that all degrees of freedom other than q are in quasi-equilibrium so that the diffusivity is a time-independent scalar that depends only on q. Note that the Bayesian scheme presented here could be applied to more complex models; however, parameter optimization could become more difficult as more parameters are added. Under the above assumptions, the Brownian integrator becomes

$$q_2 = q_1 + \beta D(q_1) f(q_1, t_1)\Delta t + (2D(q_1)\Delta t)^{1/2} g_t + \nabla D(q_1)\Delta t \tag{1}$$

where $\beta = (k_BT)^{-1}$, $\Delta t = t_2 - t_1$, $f(q_1,t_1) = f_{bias}(t_1) + f_{sys}(q_1)$ is the sum of the known time-dependent biasing force $f_{bias}(t)$ and the intrinsic system force $f_{sys}(q) = -\nabla W(q)$, and g_t is a random number taken from a Gaussian distribution with a zero mean and a variance of unity. Owing to the properties of g_t, we may express the probability for observing the system at position q_2 at time t_2 given that it was observed at position q_1 at time t_1 by

$$P[(q_2, t_2|q_1, t_1)|f, D] = \frac{1}{\sqrt{4\pi D(q_1)\Delta t}} \exp\left(-\frac{[q_2 - q_1 - \beta D(q_1) f(q_1, t_1)\Delta t - \nabla D(q_1)\Delta t]^2}{4D(q_1)\Delta t}\right) \tag{2}$$

Note that eq 2 assumes overdamped Langevin dynamics and, therefore, is not valid for systems or time scales where these conditions do not apply. In particular, we show below that analysis of MD simulations on too short of time scale yields unreliable results due to inertia and associated velocity correlation. Moreover, eq 2 relies on the flutuation-dissipation relation $D(q) = k_BT/\gamma(q)$, where $\gamma(q)$ is the position-dependent friction coefficient in the Langevin equation. Fluctuation–dissipation relations are only valid in general for systems near equilibrium; therefore, the use of eq 2 may not be justified for simulations involving nonequilibrium work methods.

Equation 2 is similar to the formula presented by Türkcan et al., with the important distinction that we have included the $\nabla D(q_1)\Delta t$ term of the Ermak and McCammon integrator, which was neglected in their formula. Unlike in Türkcan et al., we have not included the effect of noise in the measurement of the position, since we consider only simulations and not experimental measurements. Our expression also differs in that we include the contribution of the known time-dependent force $f_{bias}(t)$ to $f(q,t)$. Although we consider only a one-dimensional coordinate for simplicity, higher dimensional analogues of eq 2 can be constructed with the mapping $q \to \mathbf{r}$, $D(q) \to D(\mathbf{r})$, and $f(q) \to \mathbf{f}(\mathbf{r})$ (Türkcan et. al use a two-dimensional form).

In our implementation, $f_{sys}(q)$ and $D(q)$ are represented by piecewise cubic interpolants with continuous first derivatives. The functions are defined by their values at the nodes q_i, which have uniform spacing h. To calculate $D(q)$, we choose the node q_i nearest to q for which $q_i \le q$ and employ the formula

$$D(q) = a_i + a_i^{(1)}\frac{q - q_i}{h} + a_i^{(2)}\left(\frac{q - q_i}{h}\right)^2 + a_i^{(3)}\left(\frac{q - q_i}{h}\right)^3 \tag{3}$$

where

dx.doi.org/10.1021/ct300867e | J. Chem. Theory Comput. 2013, 9, 876–882

$$a_i^{(1)} = (-a_{i-1} + a_{i+1})/2$$
$$a_i^{(2)} = (2a_{i-1} - 5a_i + 4a_{i+1} - a_{i+2})/2$$
$$a_i^{(3)} = (-a_{i-1} + 3a_i - 3a_{i+1} + a_{i+2})/2 \quad (4)$$

To calculate $\nabla D(q)$, we use the analytic derivative of eq 3. In the domain $q \in [q_1,q_n]$, $D(q)$ is determined completely by the values $\{a_i\}$ for $i \in \{1,2,...,n\}$. For aperiodic systems, we define $a_0 = a_1$, $a_{n+1} = a_n$, and $a_{n+2} = a_n$ to maintain the validity of eq 4. For periodic systems, the values are wrapped, e.g. $a_{n+1} = a_1$. The system force $f_{sys}(q)$ is calculated in a similar manner, with the values $\{b_i\}$ parameterizing $f_{sys}(q)$ in the same way that $\{a_i\}$ parameterizes $D(q)$.

The values $\{a_i,b_i\}$ completely determine $D(q)$ and $f_{sys}(q)$; therefore, $P[(q_{\alpha+1},t_{\alpha+1}|q_\alpha,t_\alpha)|f,D]$ can be interpreted as the likelihood of observing a transition from (q_α,t_α) to $(q_{\alpha+1},t_{\alpha+1})$ under a known biasing force $f_{bias,\alpha}$ given the model parameters $\{a_i,b_i\}$. Let $\{\mathbf{d}_\alpha\} = \{q_\alpha f_{bias,\alpha}\}$ represent the trajectory data, i.e., the position and external biasing force at each frame α of the trajectory. The likelihood of the complete trajectory is the product of the probability of observing each transition:[3]

$$P(\{\mathbf{d}_\alpha\}|\{a_i, b_i\}) = \prod_{\alpha=1}^{M-1} P[(q_{\alpha+1}, t_{\alpha+1}|q_\alpha, t_\alpha, f_{bias,\alpha})|\{a_i, b_i\}] \quad (5)$$

where M is the number of points in the trajectory. By Bayes' theorem, the posterior probability of the parameters, $\{a_i,b_i\}$, given the trajectory data, $\{\mathbf{d}_\alpha\}$, is

$$P(\{a_i, b_i\}|\{\mathbf{d}_\alpha\}) \propto P(\{\mathbf{d}_\alpha\}|\{a_i, b_i\})p_{prior}(\{a_i, b_i\}) \quad (6)$$

The prior probability of the parameters, $p_{prior}(\{a_i,b_i\})$, encodes information on the appropriate distribution of parameter values as well as prior knowledge about the form of the solution. Given the role of $D(q)$ in eq 2, an appripriate prior distribution for $\{a_i\}$ is the scale invariant distribution

$$p_{scale}(\{a_i\}) = \prod_i 1/a_i \quad (7)$$

which ensures that $\ln(a_i)$ is uniformly sampled.[24]

Other prior knowledge about the parameters $\{a_i,b_i\}$ may be included by multiplying by additional prior probabilities. When a biasing method, such as ABF or metadynamics, provides an estimate of the force, f_{calc}, having an error $\delta f_{calc}(q)$, we penalize deviations from this estimate[25] by

$$p_{known}(\{b_i\}) = \prod_{i=1}^{n} \exp\left(-\frac{[f_{sys}(q_i) - f_{calc}(q_i)]^2}{2\delta f_{calc}(q)^2}\right) \quad (8)$$

Following Hummer,[3] one may also express prior knowledge that the diffusivity should be smooth on a certain length scale by

$$p_{smooth}(\{a_i\}) = \prod_{i=2}^{n} \exp\left(-\frac{[D(q_i) - D(q_{i-1})]^2}{2\varepsilon^2}\right) \quad (9)$$

We have found that usage of this prior knowledge improves results when sampling is incomplete; however, using too small a value for ε can result in a "blurred" diffusivity. For our systems, analyses using a weak penalty of $\varepsilon = 100h$ Å/ns (where h is the distance between nodes, $q_i - q_{i-1}$) provided better agreement between poorly sampled trajectories and well sampled trajectories than analyses using no penalty. For the well sampled trajectories, this penalty had no appreciable effect on the results.

To determine the Bayesian posterior distribution,[25,26] we generate a chain of states specified by particular values of $\{a_i,b_i\}_k$ employing a Monte Carlo procedure, which is outlined in Figure 1. Beginning in state k, defined by $\{a_i,b_i\}_k$, we generate

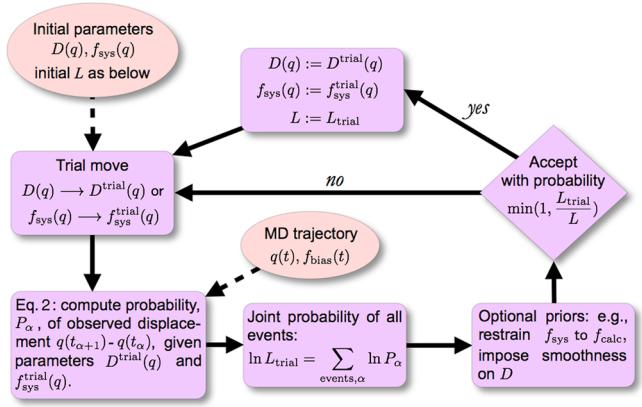


Figure 1. Flowchart depicting the Monte Carlo/Bayesian inference algorithm.

a trial state $\{a_i,b_i\}_{trial}$. An element of $\{a_i,b_i\}$ is chosen with uniform probability and modified according to $a_j' = a_j + s_D T_k$ or $b_j' = b_j + s_f T_k$, where s_D and s_f are characteristic sizes of the random steps (the values $s_D = 5$ Å^2/ns and $s_f = 0.05$ kcal/(mol Å) were used in this work) and T_k is a pseudorandom number chosen from the long-tailed[26] distribution $P(T_k) = [\pi(1 + T_k^2)]^{-1}$. The trial state $\{a_i,b_i\}_{trial}$ is accepted or rejected based on the Metropolis–Hastings criterion;[25] that is, it is accepted with the probability $\min(1,L_{trial}/L_k)$, where $L_k = P(\{\mathbf{d}_\alpha\}|\{a_i,b_i\}_k)$ $p_{prior}(\{a_i,b_i\}_k)$. p_{prior} is a product that may include one or more of p_{scale}, p_{known}, or p_{smooth}, depending on what prior knowledge is assumed. In this work, p_{scale} is always included, while p_{scale} or p_{known} are included only where explicitly stated. Figure S1 of the Supporting Information shows examples of the evolution of the acceptance ratio of the Metropolis–Hastings algorithm during the analyses, which typically stays between 0.45 and 0.65.

The optimal values $\{a_i,b_i\}_k$ are taken to be those that yield the greatest posterior probability with error bars given by the standard deviation of the Bayesian posterior distribution. An example profile of the posterior distribution is shown in Figure S2 of the Supporting Information. In practice, all computations are made with $\ln(L_k)$ to avoid overflow of the double-precision floating-point representations used by computers.

METHODS

In the present section, we describe the details of the MD simulations and the diffusivity analyses based on the Hummer Bayesian method.[3]

MD Methods. The simulated systems contained 779 water molecules, including a tagged water molecule. Interatomic interactions were calculated using the standard water model of the CHARMM[27] force field. The simulation was performed with the program NAMD.[28] A Langevin thermostat with a damping constant of 1 ps^{-1} was applied to maintain a temperature near 300 K, and a Langevin piston[29] was applied to maintain the pressure at 1 bar. The SETTLE algorithm[30] was employed to constrain O–H bond lengths and H–O–H

dx.doi.org/10.1021/ct300867e | J. Chem. Theory Comput. 2013, 9, 876–882

valence angles to their equilibrium values. The equations of motion were integrated with a 2 fs time step and multiple-timestepping,[28] where full RESPA[31] electrostatics were calculated only every 4 fs. Nonbonded energies were calculated using particle-mesh Ewald full electrostatics[32] (grid spacing 0.89 Å) and a smooth (8 to 9 Å) cutoff of the Lennard-Jones and direct Coulomb energies.

The Hummer Method. For comparison with our method, we employed the Bayesian method described by Hummer to determine position-dependent diffusivity.[3] Each state of the Monte Carlo procedure was represented by diffusivity values D_i at each bin edge i and the probability of the bin being occupied ϕ_i, which replaces the force $f_{sys}(q)$ in our method. As with our method, trial moves of the diffusivity were chosen by $D_i^{trial} = D_i + s_D T_k$, where T_k was distributed as $P(T_k) = [\pi(1 + T_k^2)]^{-1}$. Trial moves were likewise made on the values ϕ_i with equal probability, using a characteristic size $s_\phi = 0.05/n$, where n was the number of bins; ϕ_i values were normalized after each move to maintain $\sum_j^n \phi_j = 1$. We then used $\{D_i,\phi_i\}$ to construct the rate matrix $\mathbf{R}$ as prescribed by Hummer. The log-likelihood of the series of events was calculated as

$$\ln P(\{\mathbf{d}_\alpha\}|\{D_i, \phi_i\}) = \sum_{i=1}^{n}\sum_{j=1}^{n} N_{ij}[\exp(\Delta t\mathbf{R})]_{ij} \quad (10)$$

where Δt is the time interval associated with the events, N_{ij} is the number of events starting in bin j and ending in bin i, and $\exp(\mathbf{X})$ is the matrix exponential. We found that the prior proposed by Hummer reflecting the assumption that $D(q)$ is smooth (eq 9) improved the stability of the method. We chose a weak penalty of $\varepsilon = 100h$ Å/ns.

RESULTS AND DISCUSSION

Here, we describe several illustrative tests of the method presented in the Theory section, increasing realism at each step. We began with a system that is explicitly constructed to represent diffusive motion. Next, we applied our method to an MD simulation of neat water with a position-dependent diffusivity constructed by applying a thermostat that varies spatially. Finally, we demonstrate the usefulness of our method in the limit of poor sampling for a system to which the biasing method ABF is applied.

Explicitly Diffusive Model. Following Hummer,[3] we test our method by considering an explicitly diffusive process on the potential $W(q) = -k_B T \cos(2q)$ over the periodic domain $q \in [-\pi,\pi)$ Å with a position-dependent diffusivity of $D(q) = 100(2 + \sin q)$ Å^2/ns. The trajectory was generated using the Brownian integrator of eq 1 with a time step of $\Delta t = 1.0$ fs for a total time of 100 ns. No biasing method was applied ($f_{bias}(q) = 0$), and the position was recorded every 20 steps. Although eq 2 is directly related to the Brownian integrator, note that the application of our method to this trajectory is not entirely trivial since each event represents 20 iterations of the integrator.

The data shown in Figure 2 were the result of 10 000 trial moves (about ~50% of which were accepted) including and neglecting the $\nabla D(q)$ term. The accuracy of the predicted $D(q)$ was in both cases very high; however, significant distortion of $W(q)$ appeared in the analysis that neglected the $\nabla D(q)$ term. In subsequent analyses, we therefore used only eq 2, which includes the term $\nabla D\Delta t$.

Water with Position-Dependent Thermostat. For systems such as a molecule partitioning between a lipid bilayer and an aqueous phase or a molecule diffusing within a channel,

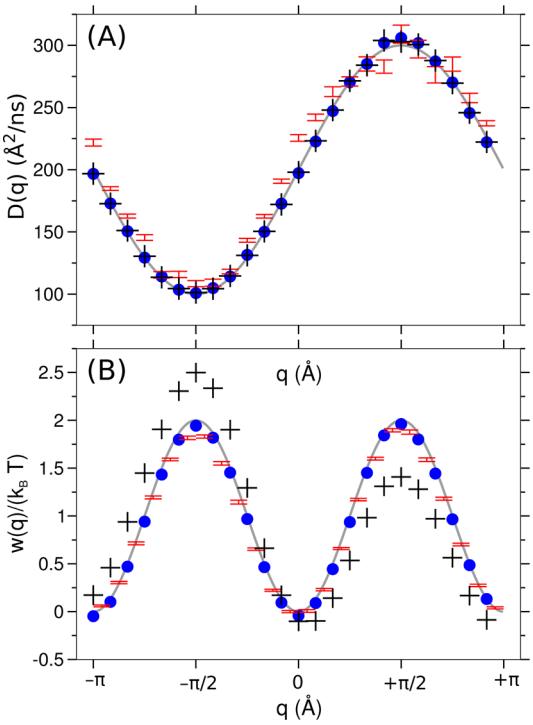


Figure 2. Calculation of the diffusivity (A) and the PMF (B) for an explicitly diffusive model. The gray line shows the exact diffusivity and PMF used in the model. Black crosses and blue circles show values determined with the $\nabla D(q)$ term of eq 2 neglected and included, respectively. In these cases, the PMF was calculated by trapezoid rule integration of the force, and the error bars (derived from the Bayesian posterior distribution) are not shown as they are significantly smaller than the symbols. The inclusion of the $\nabla D\Delta t$ term in eq 2 yields an improved estimation of the PMF. For comparison, results of the Bayesian method described by Hummer[3] are represented with red error bars.

a change in $D(q)$ usually implies a change in the local environment of the permeating molecule, which is almost invariably accompanied by a change in $W(q)$. For the purposes of testing our method, we sought to design a moderately realistic MD simulation in which the effects of $D(q)$ and $W(q)$ could be determined independently. For this reason, we employed a Langevin thermostat, which can alter system kinetics while maintaining the energetics of an isothermal ensemble.[33]

Thus, for our first test of the method using data from an MD simulation, we constructed a uniform system consisting of a cube (side length 28.5357 Å) of identical water molecules, applying a position-dependent thermostat to yield a nonuniform $D(q)$. The details of the simulation of this system are described in the Methods section. To produce a position-dependent diffusivity without affecting $W(q)$, a Langevin thermostat with a position-dependent damping parameter was applied to all molecules in the system. Specifically, the damping parameter was smoothly varied from $\gamma_0 = 0.2$ ps^{-1} for $|z| \geq 5$ Å to $\gamma_1 = 20.0$ ps^{-1} for $|z| \leq 8$ Å, using a symmetric cubic interpolation. This damping parameter is plotted on a logarithmic scale in Figure 3A.

To apply our method, we generated a 100 ns trajectory of a single "tagged" water molecule. The position of this molecule was recorded every 2 ps. The uniformity of the system implied that $W(q) = 0$; we therefore set $s_f = 0$ in the Monte Carlo/Bayesian analysis to reflect this information. The results are

dx.doi.org/10.1021/ct300867e | J. Chem. Theory Comput. 2013, 9, 876−882

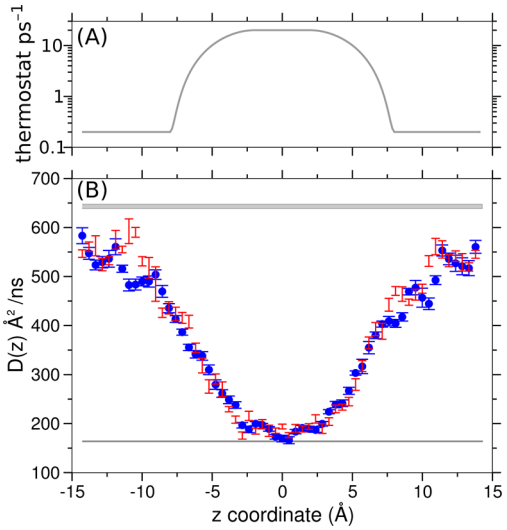


Figure 3. MD simulation of water with position-dependent thermostat. (A) Logarithmic plot of the Langevin thermostat damping parameter as a function of z. (B) Calculated diffusivity of a single water molecule as a function of z. The blue circles and associated error bars are the results of the method described here, while the red error bars are the results of the Hummer method. The gray regions represent the diffusivities calculated from the mean-squared displacement for systems having uniform damping parameters of 20.0 and 0.2 ps^{-1}. The width of these regions represents the uncertainty of the values. The interval between frames in the data was 2 ps, and the duration of the simulation was 100 ns.

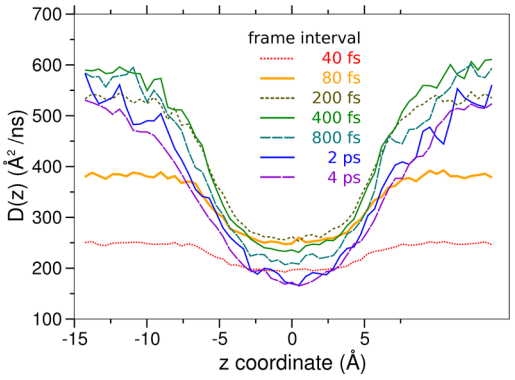


Figure 4. Dependence of the calculated diffusivity on the interval between frames. For the shortest frame intervals of 40 and 80 fs, there is significant correlation of the water molecule velocity between frames; therefore, diffusive dynamics is not a valid assumption in these cases.

presented in Figure 3B. Interestingly, $D(q)$ appears to change less abruptly than the damping parameter: the lower diffusivity in the center of the system appears to bleed beyond $|z| > 8$.

We found that our method gave results consistent with those of the Hummer method (Figure 3B). We also verified our method by comparing with the constant diffusivity of two uniform systems. We calculated the diffusivity in two similar systems with uniform Langevin damping parameters of 0.2 or 20.0 ps^{-1} using $D = \lim_{t\to\infty} \langle|\mathbf{r}(t) - \mathbf{r}(0)|^2\rangle/(6t)$,[15] with the expectation that the flat portions of Figure 3A should be similar to the uniform system. We found that the diffusivity profile was bounded by the two limiting values, namely 643 ± 4 and 163.7 ± 0.6 Å^2/ns for damping parameters of 0.2 or 20.0 ps^{-1}, respectively. However, the highest diffusity predicted by our method was about 100 Å^2/ns smaller than the value calculated for the uniform system. There is no reason to believe that the diffusivity with position-dependent damping should exactly correspond to the values for the uniform systems.

Frame Interval. To reduce simulation time and conserve disk space, trajectory information in MD simulations is typically not recorded at every step, but at fixed intervals. For example, the data used to create Figure 3B was the position of the tagged water molecule at 1000-step intervals (2 ps of simulated time). This interval ought to be kept in mind, as the results of our method (as well as the Hummer method) can depend stringently on its value. Clearly, eq 2 is only valid for displacements over which $f_{sys}(q)$ and $D(q)$ do not change considerably. One might therefore wish to reduce the frame interval (and thus the displacements) as much as possible; however, eq 2 is also not valid for short intervals where the velocity is correlated between frames. For our system, as shown in Figure 4, there is significant correlation in the velocity for frame intervals of 40 or 80 fs, which results in discrepant values for the diffusivity. Fortunately, it appears that there is a relatively wide regime of frame intervals between 200 fs and 4 ps where the diffusivity values broadly agree, notwithstanding deviations as large as 100 Å^2/ns. Hummer[3] notes that time dependence in diffusivity calculations may indicate non-Markovian dynamics and failure of the diffusive model.

Water with Position-Dependent Thermostat and External Potential. For systems in which the PMF varies substantially relative to k_BT, sampling in unbiased simulations of reasonable durations can be insufficient. Computing the PMF in such systems is made more efficient by the application of a preferential sampling method such as ABF. We hypothesized that the improved sampling provided by the biasing method could facilitate the calculation of diffusivity as well as provide a reliable estimate of the PMF, giving a complete diffusive model for the coordinate of interest. As a test of this hypothesis, we performed a simulation identical to that above (incorporating the position-dependent thermostat) except that we also applied an external potential to the tagged water molecule. This molecule was subjected to the external potential energy $V_{ext}(z) = V_0 \cos(2\pi z/\lambda)$, where V_0 = 1.5 kcal/mol, z is the z coordinate of the tagged molecule, and λ = 7.14 Å. Note that the application of this potential energy should not effect $D(q)$, which should remain that in shown in Figure 3B. Due to the fact that this potential has variations of 3 kcal/mol ≈ 5.0 k_BT, we used ABF to obtain uniform sampling and calculate the PMF.

The force applied by ABF converges[34] to $f^{\infty}_{bias} = +\nabla W(q) = -f_{sys}$; therefore, once ABF has nearly converged, the dynamics of the system is tantamount to diffusion on a flat energy surface. On a converged ABF run, we could use eq 2 with $f(q,t) = f^{\infty}_{sys} + f_{bias} = 0$ or the Hummer method quite easily. It is more difficult to use the unconverged portions of the ABF simulation. To test our method under strenuous conditions, we restrict our analysis to the first 2 ns of the trajectory where the root-mean-square deviation between the PMF calculated by ABF and the final converged PMF remains greater than 0.4 kcal/mol.

Figure 5A shows the diffusivity calculated for the ABF simulation. The error bars (given by the width of the Bayesian posterior distribution) are much larger than those depicted in Figure 3A because the ABF simulation was much shorter than the equilibrium simulation (2 ns vs 100 ns). Considering these error bars, the diffusivity calculated from the ABF simulation is in excellent agreement with that calculated from the

dx.doi.org/10.1021/ct300867e | J. Chem. Theory Comput. 2013, 9, 876−882

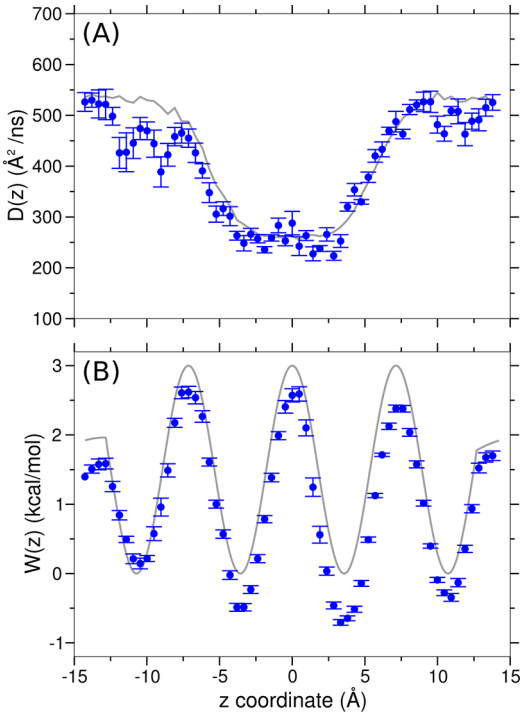


Figure 5. Calculation of the diffusivity (A) and the PMF (B) in an unconverged (2 ns) ABF simulation. The blue circles depict the diffusivity or PMF estimated by the method described herein. The simulation was similar to that used in Figure 3 except that an external potential energy (shown in gray in panel B) was applied to a single water molecule. Note that the slope of the external potential energy (and, therefore, the force) undergoes an abrupt change near $z = -12.8$ Å, which should be difficult for the method since it determines the force. For reference, the gray line in panel A is the diffusivity determined by the same method (including a frame interval of 200 fs) in a much longer equilibrium simulation where an external potential energy was not applied (see Figure 4). The adaptive biasing force method permitted this tagged water molecule to sample all parts of the system—without it, this water molecule spent most of the 2 ns in the wells of the potential.

equilibrium simulation. This was to be expected, given that the system was constructed in such a way that the diffusivity should not change upon the application of the external potential. Due to the fact that the ABF simulation was short and sampling was barely sufficient, we used assumed prior knowledge about the smoothness of $D(q)$ (see eq 9), which reduced the noisiness of the results for this simulation. A relatively short frame interval of 200 fs was chosen to maximize the number events in the Bayesian analysis. Given that the $D(q)$ error bars for the 2 ns ABF simulation are much larger than those for the 100 ns equilibrium simulation, the differences between calculations with different frame intervals (excluding 40 and 80 fs) is less significant in this case (see Figure 4).

This is the first test of the method requiring that the time-dependent biasing force, $f_{\text{bias}}(t)$, appear in the calculation. During the simulation, the force applied by the ABF algorithm was recorded with the same frequency as the z coordinate of the water molecule to give $f_{\text{bias}}(t_\alpha)$. In eq 2, $f(q_1,t_1) = f_{\text{bias}}(t_1) + f_{\text{sys}}(q_1)$, where $f_{\text{sys}}(q_1)$ is a free parameter. Because the ABF simulation was unconverged, we did not restrain $f_{\text{sys}}(q)$ to values given by ABF; i.e., we did not use eq 8. In any case, the PMF, which was calculated as $W(q) = -\int dq\, f_{\text{sys}}(q) + C$, shows modest agreement with the exact external potential, as shown in Figure 5B. Since the value of C has no physical significance, it was chosen such that the mean of the calculated PMF was the same as the mean of $V_{\text{ext}}(z)$.

Note that neither the diffusivity nor the PMF could have been calculated for this system without the application of a preferential sampling method such as ABF. In an unbiased simulation of the same system over the same time (2 ns), the tagged water molecule reached only two of the four potential wells. In fact, apart from two short transitions between the potential wells, only small regions surrounding the two potential minima were sampled. In this case, one could view ABF as a tool for obtaining the sampling necessary for an accurate diffusivity calculation in a short simulation.

CONCLUSION

Many studies which include free-energy calculations can derive a considerable benefit from the addition of kinetic information. The method presented here provides a means to calculate diffusivities and related kinetic quantities from simulations employing a wide variety of preferential sampling protocols, while adding only negligible additional computational cost (requiring only a few minutes on a contemporary multicore workstation). Thus, we envision that the method proposed here will facilitate the construction of simple, but complete, diffusive models in many future studies.

The dependence of the calculated diffusivity on the frame interval shown in Figure 4 implies deviations from the simple diffusive model. We suggest that applications of the method presented here should always include comparisons of results using different frame intervals to ensure reliability. Other checks, such as determination of the velocity autocorrelation time scale, may also be necessary.

Simple diffusive models defined by $W(q)$ and $D(q)$ provide a good approximation of dynamics for many systems; however, diffusive motion does not have a rigorous basis in general for molecular systems, and it may be necessary to move to more complex models. Fortunately, the basic Bayesian scheme adapted here, which is similar to that used by Hummer[3] and Türkcan et al.,[22] can easily accommodate different dynamical models and sources of information that have not been considered here. Thus, we anticipate that these schemes will pave the way to the use of more sophisticated models in the analysis of simulations.

For instance, hydrodynamic interactions could be incorporated by deriving the analogue of eq 2 of this work from eq 3 of Allison and McCammon,[35] which includes a tensor diffusivity. In this case, the effective hydrodynamic radius, a, of each particle might serve as a free parameter rather than $D(q)$. To incorporate velocity correlation, the analogue of eq 2 could be derived from the discretized Langevin equation[28] with the position-dependent friction coefficient, $\gamma(\mathbf{r})$, also serving as free parameters. The generalized Langevin equation[15] could also be used, with the memory kernel, $M(t)$, as an additional set of free parameters. Note, however, that including more free parameters may make it difficult for the Monte Carlo/Bayesian scheme to find unique solutions, and greater sampling may be needed than for simpler models. It has been appreciated for some time that many biomolecular systems exhibit anomalous diffusion, including telomeres in vivo,[36] proteins in crowded cytoplasm,[37] and some protein−ligand systems.[38,39] Thus, future work could be directed at developing a Bayesian scheme to parametrize models such as fractional Brownian motion or continuous-time random walks. Finally, Bayesian inference also permits

dx.doi.org/10.1021/ct300867e | J. Chem. Theory Comput. 2013, 9, 876−882

comparisons of disparate models[26] and, hence, could be used to find the best model of motion among several candidates.

■ ASSOCIATED CONTENT

Supporting Information

Included are two figures, which provide examples of the evolution of the acceptance ratio during the application of the Monte Carlo/Bayesian analysis as well as an example of the posterior distribution of one diffusivity value. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: chipot@ks.uiuc.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.C. thanks David B. Wells for kindly providing his modified version of NAMD that permits a position-dependent Langevin damping parameter. J.C. was supported by a grant from Innova-Chile CORFO (FCR-CSB 09CEII-6991). The Centro Interdisciplinario de Neurociencia de Valparaíso is a Millennium Institute (P09-022-F) supported by the Millennium Scientific Initiative of the Ministerio de Economía, Fomento y Turismo.

■ REFERENCES

(1) Kramers, H. *Physica* **1940**, *7*, 284−304.
(2) Zwanzig, R. *J. Stat. Phys.* **1983**, *30*, 255−262.
(3) Hummer, G. *New J. Phys.* **2005**, *7*, 34.
(4) Forney, M.; Janosi, L.; Kosztin, I. *Phys. Rev. E* **2008**, *78*, 051913.
(5) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer: Berlin, 2007.
(6) Lelièvre, T.; Stoltz, G.; Rousset, M. *Free Energy Computations: A Mathematical Perspective*; World Scientific: Singapore, 2010.
(7) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(8) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(9) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(10) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721−6.
(11) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120.
(12) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. *J. Chem. Theory Comput.* **2009**, *6*, 35−47.
(13) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 90601.
(14) Zwanzig, R. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2029.
(15) Mamonov, A.; Kurnikova, M.; Coalson, R. *Biophys. Chem.* **2006**, *124*, 268−278.
(16) Marrink, S.; Berendsen, H. *J. Phys. Chem.* **1994**, *98*, 4155−68.
(17) Woolf, T.; Roux, B. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 11631.
(18) Luo, Y.; Egwolf, B.; Walters, D.; Roux, B. *J. Phys. Chem. B* **2009**, 2035−2042.
(19) Comer, J.; Aksimentiev, A. *J. Phys. Chem. C* **2012**, *116*, 3376−3393.
(20) Hénin, J.; Tajkhorshid, E.; Schulten, K.; Chipot, C. *Biophys. J.* **2008**, *94*, 832−839.
(21) Wei, C.; Pohorille, A. *J. Phys. Chem. B* **2011**, *115*, 3681−3688.
(22) Türkcan, S.; Alexandrou, A.; Masson, J. *Biophys. J.* **2012**, *102*, 2288−2298.
(23) Ermak, D.; McCammon, J. *J. Chem. Phys.* **1978**, *69*, 1352.
(24) Best, R.; Hummer, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16902−16911.
(25) Dose, V. *Rep. Prog. Phys.* **2003**, *66*, 1421.
(26) von Toussaint, U. *Rev. Mod. Phys.* **2011**, *83*, 943−999.
(27) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(28) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(29) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(30) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(31) Zhou, R. H.; Harder, E.; Xu, H. F.; Berne, B. J. *J. Chem. Phys.* **2001**, 115.
(32) Batcho, P. F.; Case, D. A.; Schlick, T. *J. Chem. Phys.* **2001**, *115*, 4003−4018.
(33) Paterlini, M.; Ferguson, D. *Chem. Phys.* **1998**, *236*, 243−252.
(34) Lelièvre, T.; Rousset, M.; Stoltz, G. *Nonlinearity* **2008**, *21*, 1155.
(35) Allison, S.; McCammon, J. *Biopolymers* **1984**, *23*, 363−375.
(36) Burnecki, K.; Kepten, E.; Janczura, J.; Bronshtein, I.; Garini, Y.; Weron, A. *Biophys. J.* **2012**, *103*, 1839−1847.
(37) Weiss, M.; Elsner, M.; Kartberg, F.; Nilsson, T. *Biophys. J.* **2004**, *87*, 3518−3524.
(38) Glöckle, W.; Nonnenmacher, T. *Biophys. J.* **1995**, *68*, 46−53.
(39) Kou, S.; Xie, X. *Phys. Rev. Lett.* **2004**, *93*, 180603.