Journal of Computational Physics **151,** 190–211 (1999)
Article ID jcph.1999.6218, available online at http://www.idealibrary.com on IDEAL®

Reconstructing Potentials of Mean Force through Time Series Analysis of Steered Molecular Dynamics Simulations

Justin R. Gullingsrud, Rosemary Braun, and Klaus Schulten[1]

Department of Physics and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 North Matthews, Urbana, Illinois 61801
E-mail: justin@ks.uiuc.edu, braun@ks.uiuc.edu, kschulte@ks.uiuc.edu

Received September 1, 1998; revised January 27, 1999

Atomic force microscopy (AFM) experiments and steered molecular dynamics (SMD) simulations have revealed much about the dynamics of protein-ligand binding and unbinding, as well as the stretching and unfolding of proteins. Both techniques induce ligand unbinding or protein unfolding by applying external mechanical forces to the ligand or stretched protein. However, comparing results from these two techniques, such as the magnitude of forces required to unbind ligands, has remained a challenge since SMD simulations proceed six to nine orders of magnitude faster due to limitations in computational resources. Results of simulations and experiments can be compared through a potential of mean force (PMF). We describe and implement three time series analysis techniques for reconstructing the PMF from position and applied force data gathered from SMD trajectories. One technique, based on the WHAM theory, views the unbinding or stretching as a quasi-equilibrium process; the other two techniques, one based on van Kampen's Ω-expansion, the second on a least squares minimization of the Onsager–Machlup action with respect to the choice of PMF, assume a Langevin description of the dynamics in order to account for the nonequilibrium character of SMD data. The latter two methods are applied to SMD data taken from a simulation of the extraction of a lipid from a phospholipid membrane monolayer.

Key Words: WHAM; Langevin; Onsager–Machlup action; molecular dynamics; protein-ligand.

[1] To whom correspondence should be addressed.

0021-9991/99 $30.00

1. INTRODUCTION

Binding and unbinding processes, such as those exhibited by protein-ligand complexes, are governed by rare transitions from one equilibrium state to another. By applying an external force to the ligand, these transitions can be induced at a much faster rate than would otherwise occur. This has been realized experimentally in micromanipulation studies using atomic force microscopy (AFM) [1, 2], as well as in recent computer simulations [3–7].

In AFM experiments, the tip of an elastic cantilever applies an external mechanical force to a protein-ligand complex, for example, in such a way as to facilitate unbinding; the applied force is measured by monitoring the position of the tip. The experiments take place over time scales on the order of 1 ms to 1 s. The spring constants k of the cantilevers are typically on the order of 1 pN/Å, so that fluctuations in the position of an attached ligand, $(k_B T/k)^{1/2}$, are large on the atomic scale, e.g., 5 Å. Hence, minute details of the unbinding process are not resolved: most experiments treat the peak force, labeled the rupture force, as the primary datum.

Steered molecular dynamics (SMD) simulates a biopolymer system under the influence of an applied external force; such simulations can describe AFM-style micromanipulations in atomic-level detail and employ stiffer springs than those used in present AFM experiments. As a result, more detailed information about interaction energies as well as finer spatial resolution can be obtained. The main drawback is that, due to limitations in attainable computational speeds, simulations cover time scales that are typically 10^6 times shorter than those of AFM experiments. Several investigators [8, 9] have described means of overcoming these timescales by employing a Langevin model for the dynamics of proteins. SMD simulations have elucidated important details about the unbinding process of biotin from avidin [3, 4], retinal from bacteriorhodopsin [5], retinoic acid from its receptor [10], lipids from membranes [6, 11], as well as details about stretching of immunoglobin domains [7], fibronectin domains [12], and of other proteins [13]. A review of the SMD method is provided in [14].

There remain several challenges for SMD investigations. First, as shown in [4, 15], the rupture forces cannot be directly compared, because the unbinding in the two methods takes place in different physical regimes: SMD simulations operate in the drift regime, where the unbinding rate is limited by friction, while AFM experiments operate in a thermally activated regime. Second, a significant amount of irreversible work is generated during SMD simulations due to the high pulling speeds. This work needs to be discounted from the reversible work necessary for binding and unbinding in order to compare AFM and SMD results.

As is well known, irreversible processes connect two states A and B of a system at the expense of work $\Delta W_{A\to B}$ which exceeds the corresponding difference in thermodynamic potential ΔG_{AB}: $\Delta W_{A\to B} \geq \Delta G_{AB}$. Similarly, along an unbinding path parameterized by a coordinate x and characterized by a potential of mean force $U(x)$, the work done in going from the initial point $x = 0$ to x obeys the inequality

$$W(x) \geq U(x). \tag{1}$$

In the case of an AFM experiment one expects a protein-ligand system to remain in quasi-equilibrium, i.e., in equilibrium with respect to the overall potential due to inherent and slowly varying external forces, so that the work $W(x)$ remains close to $U(x)$. However, in

an SMD simulation $W(x)$ and $U(x)$ are expected to differ significantly. One may worry that SMD data cannot even in principle yield $U(x)$; however, this has been proven unfounded by the remarkable identity derived by Jarzynski [16, 17] which, in the present notation, reads

$$\langle \exp[-W(x)/k_BT] \rangle = \exp[-U(x)/k_BT]. \tag{2}$$

In this identity $\langle \cdots \rangle$ denotes an average over processes whose starting configurations are drawn from a canonical ensemble.

Equation (2) indicates that $U(x)$ can be reconstructed from an exponential average of the work $W(x)$. In practice, statistical error in the computed value of $U(x)$ will dominate the result unless the spread in $W(x)$ over many trajectories is less than k_BT [16]. In the trajectories examined here, the friction coefficient is large, implying due to the fluctuation-dissipation theorem that fluctuations in the work will be large compared to k_BT. Hence, Eq. (2) cannot be directly applied in these cases. Instead, by applying a model of the dynamics of the system, we seek to discount the irreversible work before any averages are made.

The new algorithms for analysis of SMD time series suggested below make two key assumptions. First, it is assumed that the relationship between a potential $U(x)$ and the system simulated by SMD is describable by means of a Langevin equation. Second, the SMD data are interpreted in terms of a model of *one-dimensional* stochastic dynamics, implying an average over the remaining degrees of freedom so that the potential $U(x)$ is a potential of mean force (PMF) [18]. Thus, from the wealth of data potentially available from an SMD simulation, the algorithms employ only two one-dimensional time series, the position $x(t)$ of the pulled ligand, and the force $F(x, t)$ exerted on it, with the addition in some cases of system-specific parameters such as the friction coefficient. Barring major structural changes to the protein induced by excessive external force, the underlying PMF should be essentially the same regardless of pulling speed.

The AFM experiments measure the force F_0 required to unbind a ligand from a protein within a certain time τ_0. In the framework of the stochastic model underlying the time series analysis suggested below, the experimental observable F_0 can be readily related to the PMF by means of the theory of mean first passage times [19]. Given $U(x)$, an applied time-independent force F, and the effective friction coefficient of the ligand one can readily evaluate the mean time $\tau(F)$ required for ligand unbinding as shown in [15]. The measured force F_0 and its distribution are then obtained from a solution of $\tau(F_0) = \tau_0$. The mean first passage time τ can also be determined for time-dependent forces [20], as they arise in AFM experiments and as they are considered below (cf. Eq. (11)). The relationship between the protein-ligand system, SMD simulation, stochastic model, time series algorithms, mean first passage time calculation, and AFM experiment is summarized in Fig. 1.

In this paper, we describe three methods for reconstructing the PMF $U(x)$ from a simulation of forced unbinding. One method is based on the Weighted Histogram Analysis Method (WHAM) [21, 18], the second is motivated by van Kampen's Ω-expansion [22], and the third involves a minimization of the Onsager–Machlup action [23] to find an optimally matching PMF for a given set of SMD data. The WHAM method assumes that the manipulated system remains in quasi-equilibrium. Even though this condition is not met in SMD simulations, we include this method since it is often applied to analyze data resulting from umbrella sampling [24, 25], a method which is closely related to SMD except that SMD accepts a high degree of irreversibility. The method related to the van Kampen

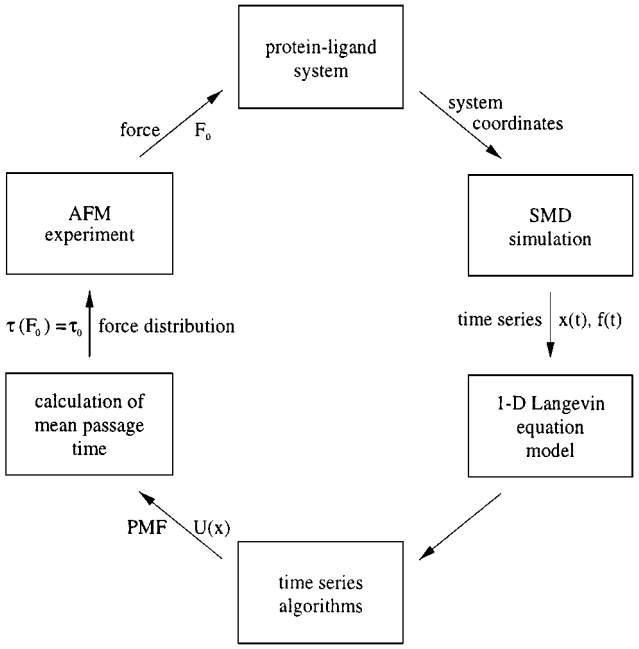

FIG. 1. Schematic representation of the relationship between the protein-ligand system, SMD simulation, stochastic model, time series algorithm, mean first passage time, and AFM experiment.

expansion is very similar to an analysis of SMD data discussed in [15]. The analysis based on the Onsager–Machlup action appears to be novel.

The three methods are applied to model systems with a single spatial dimension, but otherwise assume external forces as well as fluctuating and dissipative forces that are typical of SMD simulations of protein-ligand complexes. The model systems are described by means of a Langevin equation in the strong friction limit [22],

$$\gamma \dot{x} = F(x,t) - dU(x)/dx + \sigma\xi(t). \tag{3}$$

Here the position of the ligand is represented by the coordinate x of a very light particle, γ is the friction coefficient, assumed to be constant, $F(x,t)$ is the applied external force, $U(x)$ is the underlying potential (PMF), and $\sigma\xi(t)$ is Gaussian white noise with root mean square amplitude σ and δ-function autocorrelation, i.e., $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. σ and γ are related by the fluctuation-dissipation theorem according to $\sigma^2 = 2\gamma k_B T$, where T is the temperature and k_B is Boltzmann's constant [22].

We also demonstrate how the friction coefficient γ may be determined from an analysis of the velocity autocorrelation function of the ligand. This analysis may be performed in a separate simulation, during which the velocity of ligand atoms is sampled on a time scale equal to or faster than the velocity relaxation time m/γ. For this purpose the dynamics of the ligand must be modeled through a Langevin equation with the mass term included (cf. Eq. (29)). The force and position time series are sampled less frequently than the velocity relaxation time; hence Eq. (3) applies for the PMF analysis.

The methods for analysis of SMD data will be described in Section 2. In Section 3, the methods are applied to several types of potentials to gauge their effectiveness for different types of barrier shapes and pulling speeds. In Section 4 we apply the two nonequilibrium

methods to SMD data from lipid unbinding studies [6]. Discussion and conclusions follow in Section 5.

2. THEORY

In this section we formulate three algorithmic approaches for constructing a PMF $U(x)$ from SMD time series data, one based on an algorithm widely used for simulations near equilibrium, and two which employ nonequilibrium models for the analysis.

2.1. *The WHAM Approach*

WHAM [21, 18], an extension of umbrella sampling [24, 25], provides a means of determining an unknown underlying potential $U_0(x)$ from a set of simulations performed on a system with the addition of "biasing potentials." Each biasing potential is used to enhance the sampling in a given region, typically referred to as a window. In umbrella sampling, one performs M simulations with different biasing potentials U_i, $i = 1 \ldots M$, obtaining a probability distribution for each of the M windows. After compensating for the bias introduced by the potentials $U_i(x)$, the M distributions are "stitched" together. For this purpose, it is important that the distributions from adjacent windows overlap so that all points in x are well sampled.

In order to construct the potential $U_0(x)$, or equivalently the Boltzmann distribution $P_0(x) = Z_0^{-1} \exp[-U_0(x)/k_BT]$, WHAM combines the M measured probability distributions $P_{0i}(x)$, corresponding to potentials $U_{0i} = U_0 + U_i$, in a weighted sum such that the biasing potentials U_i are discounted. For this purpose one employs the expression

$$P_0(x) = \sum_{i=1}^{M} \omega_i(x) \exp[U_i(x)/k_BT] \frac{Z_0}{Z_{0i}} P_{0i}(x). \tag{4}$$

Here Z_0/Z_{0i} is the ratio of the configuration integral $\int dx \exp[-U_0(x)/k_BT]$ of the unbiased system to that of the system with the additional biasing potential U_i. The weighting factors $\omega_i(x)$ are subject to the normalization constraint

$$\sum_{i=1}^{M} \omega_i(x) = 1, \tag{5}$$

and are chosen such that the variance $\langle P_0^2(x)\rangle - \langle P_0(x)\rangle^2$ of $P_0(x)$ is minimized. The assumption that the M simulations are statistically independent of one another yields an expression for the weights [21, 18],

$$\omega_i(x) = \lambda(x) N_i \frac{Z_{0i}}{Z_0} \exp[-U_i(x)/k_BT], \tag{6}$$

where N_i is the number of data points taken to construct the probability distribution in simulation i and where $\lambda(x)$ is the Lagrange multiplier which serves to enforce the normalization condition (5). Substituting (6) into (4) one obtains

$$P_0(x) = \frac{\sum_{i=1}^{M} P_{0i}(x) N_i}{\sum_{i=1}^{M} (Z_0/Z_{0i}) N_i \exp[-U_i(x)/k_BT]}, \tag{7}$$

with

$$\frac{Z_0}{Z_{0i}} = \int_{x_0}^{x_f} dx \; P_0(x) \exp[-U_i(x)/k_BT]. \tag{8}$$

Equations (7), (8) can be used to obtain $P_0(x)$ and, hence, $U_0(x)$. In practice, (7) and (8) are solved iteratively until they converge.

In the case of SMD, the system is subject to an external force supplied by a pulled spring,

$$F(x,t) = k(vt - x), \tag{9}$$

where k is the spring constant and v is the velocity at which the spring is "pulled." The corresponding potential is

$$U(x,t) = \frac{1}{2}k(x - vt)^2, \tag{10}$$

which serves as the biasing potential and is parameterized by the time t rather than by the index i. We note here that applying the moving harmonic restraint (10) to a ligand moving in the potential $U_0(x)$ is equivalent to applying a time-ramped force $F(t) = kvt$ to a ligand moving in a modified potential $V(x)$,

$$-\frac{dU_0}{dx} + k(vt - x) = -\frac{dV}{dx} + kvt, \tag{11}$$

employing the definition

$$V(x) \equiv U_0(x) + \frac{1}{2}kx^2. \tag{12}$$

Thus, the analysis of [20] may be used to find the mean first passage time, once the PMF $U_0(x)$ has been determined.

Since the biasing potential (10) is altered continuously during time t rather than in M discrete steps, the time series $x(t)$ needs to be broken into M overlapping windows of time Δt. Unfortunately, this renders questionable a key assumption underlying the WHAM equations (7), (8), namely, that the data in the M windows stem from statistically independent simulations. Figure 2 illustrates how the time series data $x(t)$ are collected to form a

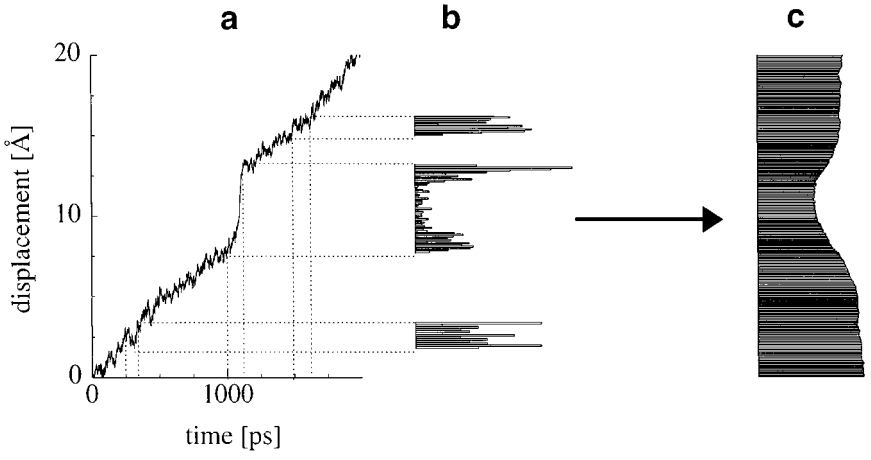

FIG. 2. Probability distribution constructed from SMD data for WHAM. (a) Displacement vs time from SMD simulation; (b) histograms $P_{0i}(x)$ of displacement values $x(t)$ for several ranges of t; (c) reconstructed probability distribution $P_0(x)$.

probability distribution $P_{0i}(x)$. In solving the WHAM equations (7), (8) for the SMD time series, we replace the factors $\exp[-U_i(x)/k_BT]$ in Eq. (7) by their time averages,

$$P_i(x) = \frac{1}{\Delta t}\int_{t_i}^{t_i+\Delta t} dt\, \exp[-U(x,t)/k_BT]. \tag{13}$$

For the biasing potential (10) the integral (13) can be evaluated, yielding

$$P_i(x) = \frac{\sqrt{\pi k_BT/k}}{2v\Delta t}\left\{\mathrm{erf}\left[\sqrt{\frac{k}{2k_BT}}(x - vt_i)\right] - \mathrm{erf}\left[\sqrt{\frac{k}{2k_BT}}(x - v(t_i+\Delta t))\right]\right\}. \tag{14}$$

2.2. *Gaussian Drift Method*

A reconstruction of the PMF $U(x)$ from an SMD time series without the assumption of quasi-equilibrium has been suggested in [15]. A particle is assumed to undergo Langevin dynamics as described by Eq. (3), with the external force (9). Approximating the velocity of the particle as $\dot{x} \approx v$ in the case of large spring constant k, and assuming the friction coefficient γ is known, $U(x)$ can be reconstructed by discounting the frictional contribution from the total work done through the integral

$$U(x) = U(0) + \int_o^x dx'(F - \gamma v). \tag{15}$$

This approach may be refined by (i) computing the velocity of the particle as a function of time from SMD data and (ii) measuring the size of the fluctuations in the the position of the particle around the deterministic trajectory, i.e., the trajectory $\tilde{x}(t)$ one would obtain from Eq. (3) in the absence of noise. It has been shown in [22] that, to lowest order in the magnitude of the fluctuations, an ensemble of Langevin trajectories will follow a Gaussian distribution $p(x,t)$ around the deterministic trajectory $\tilde{x}(t)$,

$$p(x,t) = \left(\frac{\beta}{2\pi s(t)}\right)^{1/2}\exp\left[-\frac{1}{2\,s(t)}\beta\,(x - \tilde{x}(t))^2\right], \tag{16}$$

with $\beta = 2\gamma^2/\sigma^2$ and with a variance $s(t)$ obeying the time evolution

$$\dot{s}(t) = 2[A(t)\,s(t) + 1], \qquad A(t) \equiv \frac{d}{dx}\left[F(x,t) - \frac{dU}{dx}\right]\Bigg|_{\tilde{x}}. \tag{17}$$

When applied to SMD simulations, Eqs. (16), (17) provide a means, which we term the Gaussian drift method, for reconstructing $U(x)$ based on estimates of the deterministic trajectory and the size of the fluctuations around the deterministic trajectory. In fact, Eqs. (16) and (17) relate the first and second derivatives, respectively, of the PMF to the known external force and the deterministic trajectory. Equation (16) implies that slightly noisy trajectories will fluctuate about the deterministic trajectory, so that

$$U'(\tilde{x}(t)) = F(\tilde{x}(t), t) - \gamma\dot{\tilde{x}}(t). \tag{18}$$

The equation of motion for the width of the Gaussian relates the second derivative of the potential to information from the trajectory according to

$$U''(\tilde{x}(t)) = -\frac{\gamma}{2}\frac{\dot{s}(t)}{s(t)} + \frac{dF}{dx} + \frac{\gamma}{s(t)}. \tag{19}$$

$U(x)$ can be obtained from Eqs. (18), (19) by integration, resulting in

$$U(\tilde{x}_{i+1}) = (\tilde{x}_{i+1} - \tilde{x}_i)U'(\tilde{x}_i) + \frac{1}{2}(\tilde{x}_{i+1} - \tilde{x}_i)^2 U''(\tilde{x}_i). \tag{20}$$

If the second term in (20) is neglected and the velocity assumed constant, the method outlined reduces to that suggested in [15]. Equation (20) assumes that a unique point $\tilde{x}_i$ can be assigned to each time point t_i, which requires that $\tilde{x}(t)$ is monotonic in t, and hence one-to-one. Since the particle is continuously pulled in one direction by a relatively stiff spring, $x(t)$ can be rendered monotonic by smoothing.

2.3. *Onsager–Machlup Action Minimization Methods*

The trajectories obtained from Eq. (3) are stochastic and depend on the particular realization of the Gaussian random variable $\xi(t)$. Hence, given a set of N space-time points (x_i, t_i), there exists some probability for a Langevin trajectory of N steps to pass within Δx of these points. If there are no fluctuations ($\sigma = 0$), then the probability would vanish for all sets of (x_i, t_i) except for the deterministic trajectory determined by the force terms. For nonzero fluctuations, the probability $p(x(t)|U(x))$ of a path $x(t)$ is proportional to $\exp[-S_{OM}]$, where S_{OM}, the Onsager–Machlup action [23, 26], is the following functional of $x(t)$:

$$S_{OM} \propto \frac{1}{2}\int dt \left\{ [\gamma\dot{x} - F(x,t) + dU/dx]^2 - k_B T \frac{d^2U}{dx^2} \right\}. \tag{21}$$

The second derivative term d^2U/dx^2 arises from the Jacobian of the transformation from "Wiener coordinates" ($\xi(\cdot)$) [26] to spatial coordinates ($q(\cdot)$). Heuristically, the distance between successive points in a Wiener process is Gaussian distributed, but the distance between successive points in a Langevin trajectory can shrink or grow, depending on the shape of the potential. Physically, a nonzero curvature in the potential causes a spreading or narrowing of the distribution of points at a given time in an ensemble of Langevin trajectories.

The most likely path, given the form of $F(x, t)$ and $U(x)$, corresponds to the trajectory $x(t)$ that minimizes S_{OM}. Our interest, however, is in reconstructing $U(x)$, given a trajectory. Rather than minimize S_{OM} to find the most probable path, we assume that the most likely potential for a given time series $x(t)$ can be determined by minimizing S_{OM} with respect to the choice of the potential. In order to obtain such a choice we expand the force $-dU/dx$ in terms of a set of basis functions $f_n(x)$ to be specified further below:

$$-\frac{dU}{dx} = \sum_{n=1}^{n_{max}} c_n f_n(x). \tag{22}$$

The minimization problem will be solved with respect to the expansion coefficients $c \equiv (c_1, c_2, \ldots c_{n_{max}})^T$. To simplify the notation we define $f_o(t) \equiv -\dot{x}$, $c_o \equiv \gamma$, $\tilde{f}_n(t) \equiv f_n(x(t))$, $\tilde{F}(t) \equiv F(x(t), t)$ and rewrite Eq. (21) as

$$S_{OM} \propto \frac{1}{2}\int_{t_o}^{t_f} dt \left\{ \left[\sum_{n=0}^{n_{max}} c_n \tilde{f}_n(t) + \tilde{F}(t) \right]^2 + k_B T \sum_{n=1}^{n_{max}} c_n \left. \frac{df_n}{dx} \right|_{x(t)} \right\}. \tag{23}$$

Numerical integration of (23) yields a double sum from the square term and a single sum from the cross terms and the derivative terms; the term equal to the integral of the square of the applied force can be discarded since it does not depend on c. The resulting expression can be put into the convenient matrix form

$$S(c) = \frac{1}{2} c^T H c - b^T c, \tag{24}$$

where the definitions

$$\begin{aligned} H_{mn} &= \int_{t_o}^{t_f} dt\, \tilde{f}_m(t)\, \tilde{f}_n(t), \\ b_n &= -\int_{t_o}^{t_f} dt \left[\tilde{f}_n(t) \tilde{F}(t) + \frac{1}{2} k_B T \left. \frac{df_n}{dx} \right|_{x(t)} \right], \qquad n \geq 1, \\ b_0 &= -\int_{t_o}^{t_f} dt\, \tilde{f}_o(t) \tilde{F}(t) \end{aligned} \tag{25}$$

have been used. Expression (24) describes a generalized quadratic form with a critical point at c^* satisfying

$$Hc^* = b. \tag{26}$$

If the Hessian matrix H is positive definite, then c^* corresponds to the global minimum of S_{OM} in the chosen parameter space. If the number of basis functions n_{max} is large, H is likely to be nearly singular which requires caution in solving (26). Either n_{max} must be kept small enough to maintain numerical stability, or numerical linear algebra methods such as singular value decomposition must be employed to construct the pseudo-inverse of H and avoid divergences.

By neglecting the second derivative term, the minimization problem can be recast into a linear least squares problem consisting of one linear equation in the components of c for each time step. A least squares solution of this system of equations is more demanding in terms of memory and computational requirements than the solution of (26), but tends to be more stable when large numbers of basis functions are used (n_{max} large in Eq. (22)).

To derive the least squares method, we note that the data points are sampled evenly in time; thus the integral may be evaluated as the sum of the squares of the integrand values,

$$S_{OM} \propto \sum_{i=1}^{i_{max}} \left[\sum_{n=0}^{n_{max}} c_n \tilde{f}_n(t_i) + \tilde{F}(t_i) \right]^2. \tag{27}$$

Since the stochastic term is Gaussian, the terms in the sum should be normally distributed about zero. It follows from this that the most probable PMF is precisely the one that minimizes the sum of the squares of the terms. If the expression for the force $-dU/dx$ is nonlinear in the parameters $\{c_n\}$, gradient search algorithms or other means must be employed to find the global minimum of the action. We have only used linear models in order to reduce computation time and maximize reliability and robustness; nonlinear methods can be much slower and often require a good initial guess for the coefficient vector c.

Given a particular PMF $U(x)$, stochastic trajectories obtained by integrating Eq. (3) will tend to be found in the *neighborhood* of the most probable trajectory. Vice versa, to each individual trajectory corresponds a slightly different potential which minimizes the Onsager–Machlup action for that trajectory. Combining results from multiple trajectories helps to distinguish spurious statistical fluctuations in a single reconstruction from reproducible details of the underlying PMF. To perform the analysis for a set of stochastic trajectories $\{x(t)\}$, we seek the single PMF that maximizes the probability of the occurrence of all the trajectories in the set. Denoting by $p(U|x_i(t))$ the probability for trajectory $x_i(t)$ to occur for $U(x)$, the probability that all trajectories occur for a given $U(x)$ is

$$\begin{aligned} p(U|\{x_i(t)\}) &= p\,(U|x_1(t)) \times p\,(U|x_2(t)) \times \cdots \\ &\propto \exp(-S[x_1(t)]) \times \exp(-S[x_2(t)]) \times \cdots \\ &= \exp\left(-\sum_i S[x_i(t)]\right). \end{aligned} \tag{28}$$

Thus, multiple trajectories may be incorporated into the determination of $U(x)$ by minimizing the sum of the Onsager–Machlup functions of all individual trajectories. For S given by Eq. (24) one can determine $U(x)$ by summing the Hessian matrix H and the vector b in Eq. (25) of each trajectory before solving Eq. (26). Alternatively, multiple trajectories can be incorporated into the least squares method simply by adding the respective Onsager-Machlup actions in Eq. (27). However, the resulting system of equations can be quite unwieldly.

Despite the importance, demonstrated below, of multiple trajectories for obtaining accurate PMFs, results from the analysis of single trajectories can also be helpful in comparing the effectiveness of the reconstruction methods themselves. Single trajectory results are also of interest in view of the fact that SMD simulations of biopolymers are often extremely costly, such that only a few trajectories are available for analysis.

2.4. *Velocity Autocorrelation Analysis*

In order to discount the irreversible work generated by pulling particles along a reaction path in the framework of the Langevin model as stated by Eq. (3), one needs to know the friction coefficient γ. This coefficient can be determined through the particle's velocity relaxation, which proceeds typically in a picosecond or less. The displacement of the particle on this time scale is governed by a Langevin equation which includes the inertia term

$$m\ddot{x} = -\gamma\dot{x} + F(x,t) - dU(x)/dx + \sigma\xi(t), \tag{29}$$

where m is the mass of the particle, and γ and σ are as defined in Eq. (3). We assume in the following that the potential $U(x)$ makes a negligible contribution to the velocity correlation function; this assumption will be shown below to hold for the strongly overdamped regimes we are considering. For $F(x,t)$ as given by Eq. (9), one introduces the new variable $y \equiv x - vt + \gamma v/k$ and obtains

$$m\ddot{y} = -\gamma\dot{y} - ky + \sigma\xi(t). \tag{30}$$

This equation describes the motion of a Brownian particle in a harmonic well. Setting $\dot{y} \equiv w$, Eq. (29) reads $m\dot{w} = -\gamma w - ky + \sigma\xi(t)$. Hence, a multivariate Ornstein-Uhlenbeck process [27] describes the evolution of the variable $\mathbf{y} \equiv (y,\ w)$.

The autocorrelation function $C_w(t)$ of the velocity w, normalized by the variance, is defined as

$$C_w(t) \equiv \frac{\langle [w(t'+t) - \langle w \rangle][w(t') - \langle w \rangle] \rangle}{\langle [w(t') - \langle w \rangle]^2 \rangle}. \tag{31}$$

Following [27], for the system described by Eq. (30) one obtains the velocity autocorrelation function

$$C(t) = \frac{1}{2} e^{-\gamma t/2m} \left[\left(1 + \frac{\gamma}{\nu m} \right) e^{-\nu t/2} + \left(1 - \frac{\gamma}{\nu m} \right) e^{\nu t/2} \right], \tag{32}$$

where

$$\nu \equiv \left[\left(\frac{\gamma}{m} \right)^2 - 4k/m \right]^{1/2}. \tag{33}$$

Equation (32) corresponds to an overdamped oscillator. If the spring constant k is made sufficiently stiff, ν becomes imaginary and one obtains an alternate (though equivalent) expression for $C(t)$,

$$C(t) = e^{-\gamma t/2m} \left[\cos \frac{\omega t}{2} - \frac{\gamma}{m\omega} \sin \frac{\omega t}{2} \right], \tag{34}$$

where

$$\omega \equiv \left[4k/m - \left(\frac{\gamma}{m} \right)^2 \right]^{1/2}. \tag{35}$$

Since both k and m are known, the velocity autocorrelation may be fitted to either Eq. (32) or Eq. (34) to find γ.

3. APPLICATION TO MODEL SYSTEMS

In order to gauge the accuracy of the three reconstruction methods outlined in Section 2, we performed simulations providing solutions of the Langevin equation (3) for known one-dimensional potentials and the external force (9) and used the resulting displacement and force time series to reconstruct the respective PMFs. Since in this case the time series stemmed from a truly one-dimensional system conforming exactly to the model underlying the nonequilibrium analysis methods, applications of the three methods suggested in Section 2 should reproduce the original potentials, except for the shortcomings of the methods, mainly due to limited sampling. The test calculations were carried out for three generic types of potential $U(x)$: a sinusoidal potential, Gaussian barriers of different heights, and step function potentials of different widths. The details of the potentials and the pulling velocities chosen are provided in Table I. Except for the shape of the potential and the pulling velocity, all other model parameters were fixed: $\gamma = 4000$ pN $\cdot$ ps/Å, $k = 300$ pN/Å, and $\sigma = (2\gamma k_B T)^{1/2}$ with $T = 300$ K. For meaningful comparison, the three reconstruction methods were applied to the same trajectory data for all simulation conditions.

TABLE I
Simulation Conditions for Model Systems

Potential	Height (kcal/mol)	Width (Å)	Velocity (Å/ps)
Sinusoid	30	10	0.01, 0.001
Gaussian	10, 30, 60	5 (FWHM)	0.01
Step	30	0.5, 1.0, 4.0	0.01

3.1. *Reconstruction Parameters*

Time series of displacement and force for selected trajectories are shown in Fig. 3. Since the potentials employed in the tests have approximately the same change in amplitude from peak to trough, the same width over which the potential rises, and were generated with identical pulling conditions of velocity, spring constant, etc., the time series for $x(t)$ look very similar: one can discern in all three a jump in the displacement of similar size and at similar times, corresponding to the crossing of the top of the barriers. It appears extremely challenging to devise a reconstruction method that can distinguish between these three similar looking cases and to identify properly the sinusoidal, Gaussian, and step function potentials. However, the associated force values, related to the displacement values through Eq. (9), exhibit characteristics which better reflect the underlying potentials. For example, even though the three force time series exhibit a change in value of approximately the same

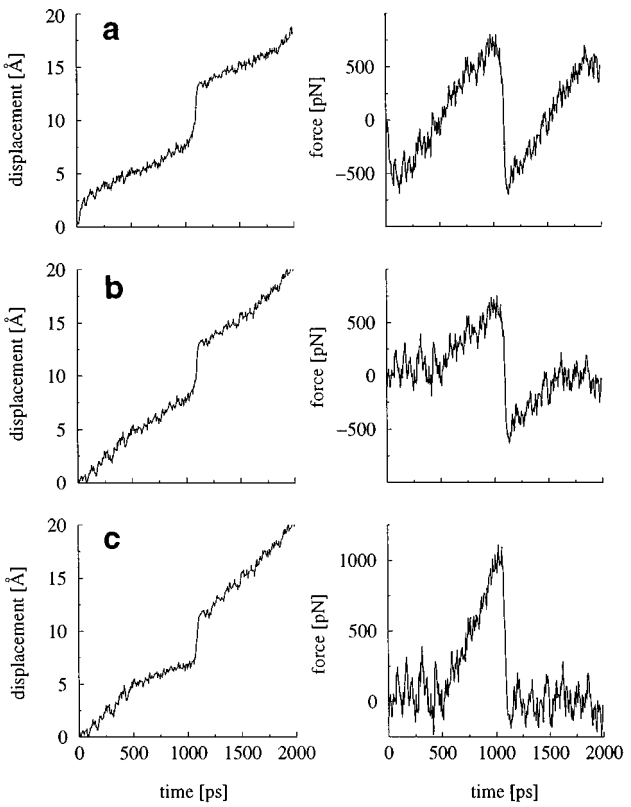

FIG. 3. Position and force time series from SMD simulation of three types of one-dimensional potentials. (a) Sinusoidal potential, $v = 10^{-2}$ Å/ps; (b) Gaussian potential, height 30 kcal/mol; (c) step function, width $\lambda = 2.0$ Å.

size, one can recognize that in the case of the sinusoidal and Gaussian potentials the force values become negative shortly after the respective systems have been pulled over the potential barriers, i.e., in the space interval where the systems experience a strong negative potential slope pushing the systems towards the pulling direction. Such negative applied forces do not occur in the case of the step function potential, where the barrier is followed by an interval of zero slope.

For the WHAM method, the trajectory was divided into 20 windows of width 125 ps for the $v = 0.01$ Å/ps simulations and 60 windows of width 360 ps for the $v = 0.001$ Å/ps simulation. The integration of (8) was performed as a Riemann sum of 200 points, and Eqs. (7), (8) were solved iteratively until the fractional change in all values of (8) (one value per window) was less than 0.005.

For the Gaussian drift method, the trajectory $x(t)$ was smoothed to monotonicity to obtain $\tilde{x}(t)$, using Savitzky–Golay filters as described in Press *et al.* [28]. Different window sizes for different trajectories were used to avoid over-smoothing the jump in $x(t)$ (see Fig. 3) where the particle crosses the peak of the potential. The velocity $\dot{\tilde{x}}$, mean square fluctuation $\tilde{s}$, and $\dot{\tilde{s}}$, were also computed in this manner. The second derivative of the potential, as reconstructed from Eq. (19), was recognizable in most of the simulations as being reasonably accurate, though somewhat noisier than the first derivative data obtained from Eq. (18). However, since the potential was very finely sampled for all the test simulations, the second derivative term made a negligible contribution to the reconstructed potential.

For the least squares method, based on Eq. (27), no smoothing of x or f was done. The velocity was computed in all cases with a 3rd order Savitzky–Golay filter corresponding to a spatial window of 0.6 Å. The basis functions were the trigonometric functions $\sin(nx/L)$ and $\cos(nx/L)$, $n = 1 \ldots 10$, for a total of 20 basis functions. The trajectory $x(t)$ was scaled and shifted so that x/L lay in the range $[-\pi, \pi]$. Such an arrangement effectively clamps the value of the potential at the final displacement to zero, since the definite integral of each of the basis functions in this range is zero. For the reconstruction of the step functions, the final value was clamped to 30 kcal/mol by subtracting a constant offset from the force time series, then adding it back after the best fit parameters were obtained.

The model potentials were also reconstructed using the full Onsager–Machlup action, based on Eq. (26). Two sets of reconstructions were done, both of which employed a basis set of 12 trigonometric functions as above, with the difference that the trajectory was scaled to $[0, \pi]$ rather than $[-\pi, \pi]$; as a result, there is no bias in the final value of the potential. In the first set of reconstructions, however, the friction coefficient was treated as an unknown parameter, while in the second set the friction coefficient was clamped to the correct value. In these two sets of reconstructions we also examine the effect of combining many independent trajectories; for this purpose 20 trajectories were generated and analyzed.

For the Gaussian drift and action minimization methods, the MATLAB computing environment and its built-in linear algebra and minimization routines were used.

3.2. *PMF Analysis*

Results of the WHAM, Gaussian drift, and least squares analyses are shown in Figs. 4–6. Results of the full action minimization analyses are shown in Figs. 7 and 8.

The Gaussian drift and least squares reconstructions show good agreement with the sinusoidal potential for both pulling speeds. This is attributable to the smoothness of the potential: higher order derivative contributions to the Gaussian drift reconstruction are negligible, and the potential is well approximated by relatively few basis functions in the least

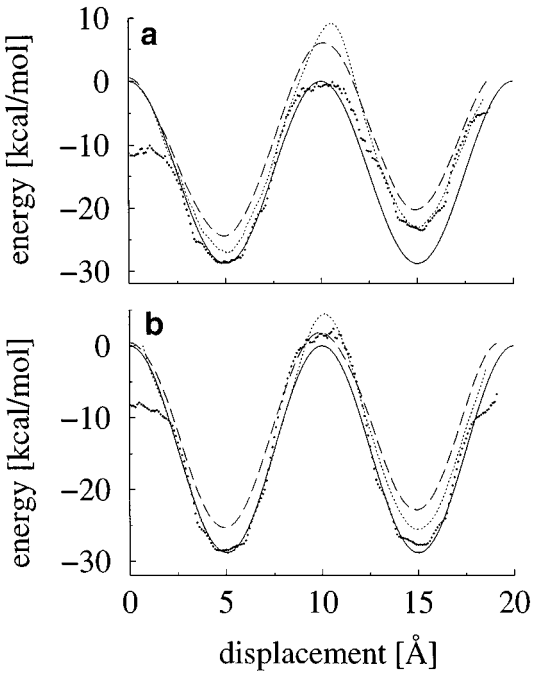

FIG. 4. Reconstruction of sinusoidal potentials using three different analysis methods, compared to the actual potential. (a) $v = 10^{-2}$ Å/ps; (b) $v = 10^{-3}$ Å/ps. Large dots, WHAM; small dots, Gaussian drift method; long-dashed line, least squares method; solid line, actual potential $U(x)$.

squares method. The lack of improvement in the reconstructions from these two methods at slower pulling speeds means that the potential is already adequately sampled at the faster speed. Under these conditions, the velocity does not induce an overall tilt in the reconstruction, in contrast to WHAM, which improves with slower pulling speeds even when the potential is well-sampled.

Reconstructions of the Gaussian potential show the effect of varying barrier height. For the 10 kcal/mol barrier (Fig. 5a), the force data are considerably obscured by temperature-induced fluctuations. The Gaussian drift reconstruction, though quite noisy, still recovers the correct height of the barrier. The least squares method fits the potential quite well, up to an overall shift; as with the sinusoidal potential, the barrier is short and smooth, easily suited to representation by a few basis functions.

For the 60 kcal/mol barrier (Fig. 5c), the particle has been pulled too fast to follow the downward slope of the potential. The force data are simply insufficient to determine the shape of the potential in this region, as the spring forces remain large and positive even after the particle is over the hill. The point at which the spring force changes sign (for examples, see Fig. 3) occurs when $x(t) \approx 10$ Å for the 10 kcal/mol barrier, 11.1 Å for the 30 kcal/mol barrier, and 12.5 Å for the 60 kcal/mol barrier. For the least squares method, since the final value of the potential is clamped to zero while the amount of irreversible work is underestimated, the reconstruction must make a compromise; the compromise is made in the region where the least amount of trajectory data is available, which corresponds to the the peak of the potential. In the WHAM reconstruction, the steep slope results in poor sampling leading to non-overlapping probability distributions for the M windows. The gaps in the distributions are manifested as discontinuities in the reconstruction, as can be seen in Figs. 5c and 6a.

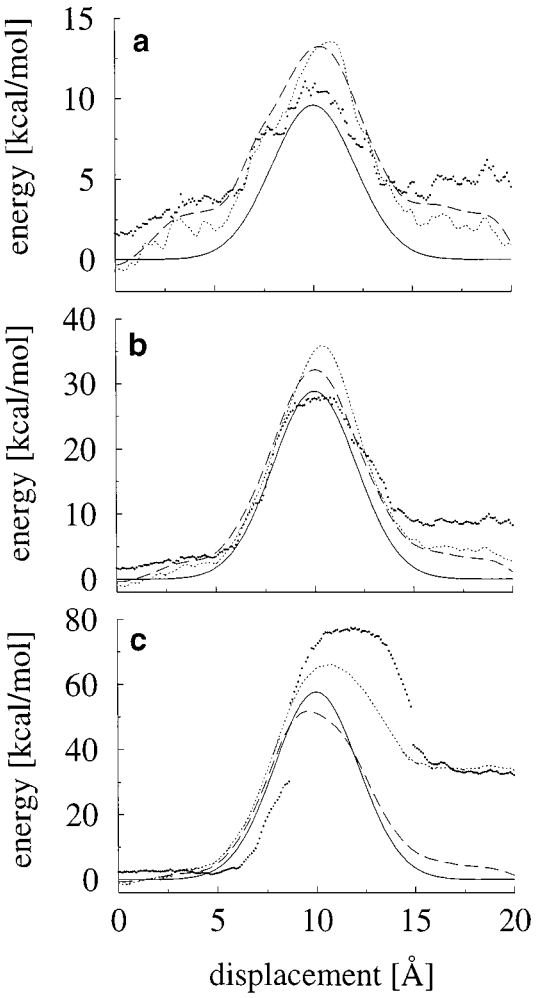

FIG. 5. Reconstruction of Gaussian potentials, of the form $U(x) = U_0 \exp[-(x - 10)^2/9]$. (a) $U_0 = 10$ kcal/mol; (b) $U_0 = 30$ kcal/mol; (c) $U_0 = 60$ kcal/mol. In all cases, $v = 10^{-2}$ Å/ps. Large dots, WHAM; small dots, Gaussian drift method; long-dashed line, least squares method; solid line, actual potential $U(x)$.

Finally, for the 30 kcal/mol barrier (Fig. 5b), the shape of the potential, the velocity, and the spring constant all contribute to make a good reconstruction possible. The measured spring forces are sufficiently large so as not to be obscured by noise, as in the small barrier. Moreover, the pulling velocity was sufficiently slow, and the spring was sufficiently stiff, to yield a clear picture of the shape of the downhill side of the potential.

Both the Gaussian drift and the least squares method were quite successful in reconstructing the step functions of Fig. 6. It can be seen that, as for the case of the Gaussian barriers, the reconstructions were especially accurate on the uphill slope of the potential, with the exception of the WHAM method for the steep $\lambda = 1$ Å step. Near the top of the slope, once again, all methods exhibit a tendency to overshoot the actual height of the step. Away from this region, the error in the Gaussian drift reconstruction can most likely be attributed to random deviations, as described in [15] and illustrated here.

In performing reconstructions using multiple trajectories, we observed that the result obtained by summing the actions, as described at the end of Section 2, was nearly identical

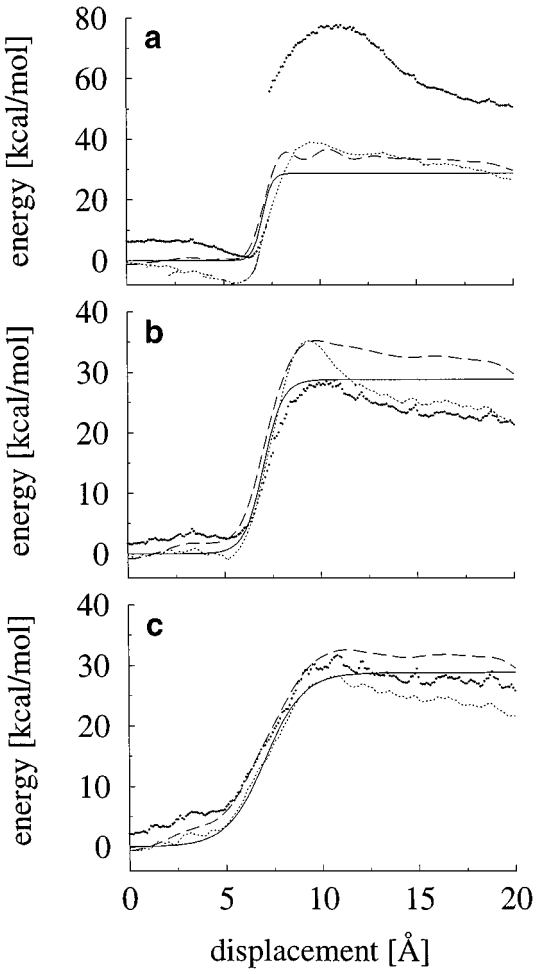

FIG. 6. Reconstruction of step function potentials of the form $U(x) = \frac{1}{2}[1 + \tanh(2(x-7)/\lambda)]$, where $U_0 = 30$ kcal/mol. (a) $\lambda = 1$ Å; (b) $\lambda = 2$ Å; (c) $\lambda = 4$ Å. Large dots, WHAM; small dots, Gaussian drift method; long-dashed line, least squares method; solid line, actual potential $U(x)$.

to the average of the potentials obtained from single trajectories. An estimate of the statistical error can be made based on the distribution of values of the potential at each displacement x. Figure 7 shows the variance σ_U^2 as a function of displacement for trajectories made with a flat potential $U(x) \equiv 0$, and with a Gaussian potential corresponding to Fig. 5b. The variance was calculated from 100 trajectories in each case. We show for comparison the theoretical estimate for the variance of a reconstructed potential derived in [15]; this estimate depends on the actual potential being relatively flat in comparison to the spring constant k. There is good agreement between the theoretical estimate and the computed value in the case of the flat potential. For the Gaussian barrier, deviations from the theoretical estimate are found in the region just past the peak of the barrier, where the curvature of the potential as well as the speed of the particle are large. In both cases, the actual potential $U(x)$ falls within the statistical uncertainty $(\sigma_U^2/N)^{1/2}$, $N = 20$, of the mean value of the reconstructed potentials.

Figure 8 shows the result of applying the action minimization methods using the full Onsager–Machlup action to the time series data of Fig. 3, using both a clamped and

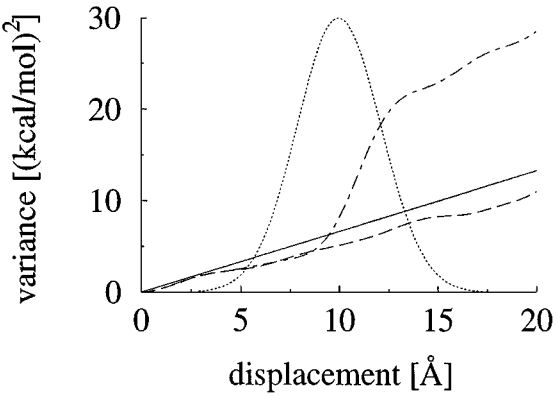

FIG. 7. Variance of potentials reconstructed from 100 single trajectories, using a flat potential $U(x) \equiv 0$ and a Gaussian barrier corresponding to Fig. 5b. Dashed line, variance from flat potential; dot-dashed line, variance from Gaussian barrier; solid line, variance based on theoretical estimate from [15], $\sigma_U^2(x) = 2k_B T \gamma v x$; dotted line, potential corresponding to Gaussian barrier, in arbitrary units of height.

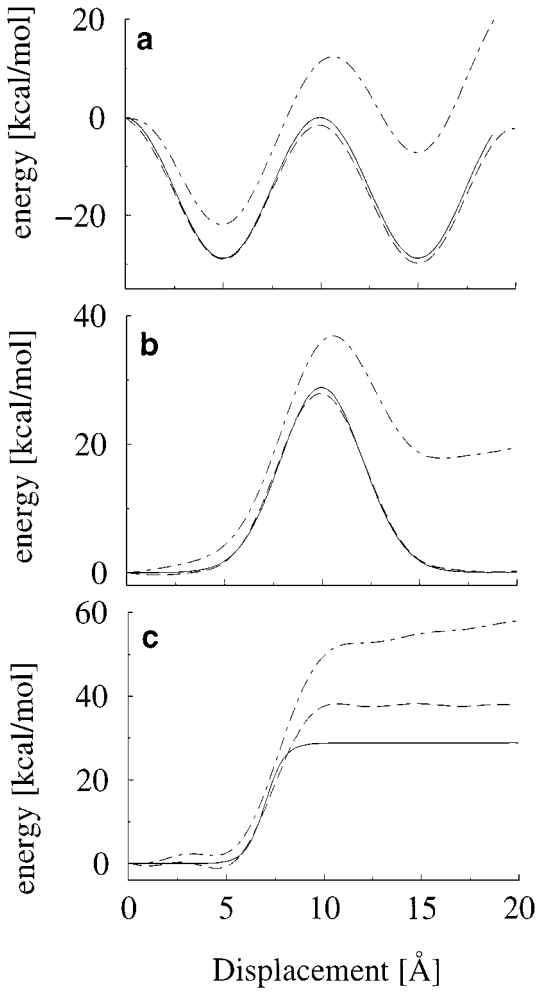

FIG. 8. Application of action minimization methods to multiple trajectories, using the three potentials of Fig. 3. (a) Sinusoidal potential, $v = 10^{-2}$ Å/ps; (b) Gaussian potential, height 30 kcal/mol; (c) step function, width $\lambda = 2.0$ Å. Dot-dashed line, unclamped friction coefficient; long-dashed line, clamped friction coefficient; solid line, actual potential $U(x)$. Results shown are averages over 20 trajectories.

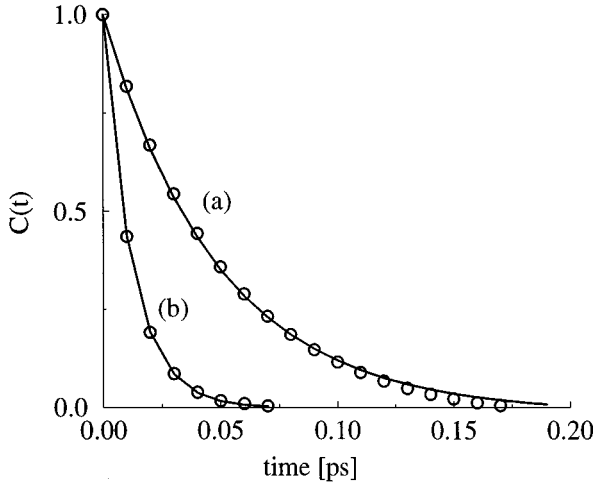

FIG. 9. Velocity autocorrelation function $C(t)$. Circles, values computed from integration of Langevin equation (29); lines, least squares fits of expression (32). (a) $\gamma = 1000$ pN · ps/Å; (b) $\gamma = 4000$ pN · ps/Å.

unclamped friction coefficient. Bearing in mind that the unclamped implementation makes no assumptions about the final value of the potential or the friction coefficient, it is not surprising that the method does not follow the shape of the potential as well as either the Gaussian drift or the clamped least squares methods. The fact that the irreversible work is underestimated in each case can be understood from the form of the Onsager–Machlup function used in the fit. Since the friction coefficient γ is an adjustable parameter, one way for the action to be minimized is simply to set γ to 0 and equate the derivative of the PMF with the spring force $F(x(t), t)$; as seen in the Gaussian drift method, the second derivative of the potential makes only a small contribution.

When the friction coefficient is clamped to the correct value, quite accurate reconstructions are obtained, except in the case of the step function. In that case, it is likely that limited sampling or imprecise calculation of the velocity where the particle reaches the top of the step leads to an underestimate of the amount of irreversible work done in this region. It is also possible that the basis functions, which are nonlocal, are inadequate for reproducing the sharp step when the rest of the potential is entirely flat; however, this does not explain why, on average, the reconstructed potential overestimates the height of the step.

3.3. *Analysis of Friction Coefficient*

Figure 8 demonstrates that knowledge of the friction coefficients is a prerequisite for accurate reconstruction of PMFs. We want to demonstrate that an analysis of velocity relaxation can yield an accurate friction coefficient. For this purpose, two sets of simulations were performed, one for $\gamma = 1000$ pN · ps/Å, and one for $\gamma = 4000$ pN · ps/Å. The motion was described by Eq. (29), with the mass set to $m = 300$ daltons, approximately the mass of biotin. Velocities calculated using the integration algorithm of [29, 30] with a time step of 1 fs were saved every 10 fs. Calculations were carried out for the sinusoidal potential as introduced above, with a pulling speed $v = 0.01$ Å/ps. The velocity autocorrelation function $C(t)$ was determined, and the friction coefficient was calculated through a least squares fit of Eq. (32) to the autocorrelation function.

The velocity autocorrelation function $C(t)$ and the fits of the analytical form (32) for the two fitted values of γ are shown in Fig. 9. The fitted friction coefficients were 4137.3 and 1034.2, exhibiting a relative error of 3% in each case.

4. TIME SERIES ANALYSIS FOR AN SMD SIMULATION PULLING A LIPID FROM A MEMBRANE

The Gaussian drift and full action minimization methods were also applied to actual SMD simulation data from a phospholipid membrane monolayer system. The system and simulation parameters are described in [6]. The lipid was pulled by a spring with velocity $v = 0.014$ Å/ps and spring constant $k = 700$ pN/Å. The velocity of the lipid was obtained from $x(t)$ with a 3rd order Savitzky–Golay filter of spatial width 0.85 Å; the spring force time series was smoothed to the same degree. A friction constant of $\gamma = 4000$ pN $\cdot$ ps/Å was assumed for both methods, which compares well with the value obtained in [11] (5000 pN $\cdot$ ps/Å). For the action minimization method, 50 trigonometric functions were used to parameterize the force.

The force used in the reconstructions and the PMF reconstructed using the two methods are shown in Fig.10. The reconstructions exhibit several important features. First, the peaks in the applied force (top graph) at $x = 3.1$, 5.6, and 8.5 Å line up with the uphill slopes of both PMF reconstructions. Second, the height of the barrier at the first peak is approximately 1.5 kcal/mol, an appropriate height value for the breaking of a single hydrogen bond [6]. Third, the action minimization reconstruction is essentially identical to the Gaussian drift reconstruction. Apparently, the basis functions used are able to reconstruct multiple features in a single trajectory and place these features in their correct spatial positions without bias.

5. DISCUSSION

In this paper, three methods for reconstructing one-dimensional potentials of mean force from the trajectory and force data gathered in SMD simulations have been described. The aim has been to develop tools that can be applied to SMD simulations as they are practiced; thus, simulation conditions have been chosen that are representative of many full-scale protein-ligand simulations, in order to test the effectiveness of the methods in picking out the features in the potential of the type that one might expect to find.

Some difficulties are encountered in the application of WHAM to analyze SMD data. First, since equilibrium is assumed, the irreversible work is taken to be zero, and thus not properly discounted; this error is inherent to the WHAM formulation. Second, it has been seen that for steep potentials the sampling is not sufficient in the region of the greatest slope to yield overlapping distributions. This difficulty may be addressed by performing additional SMD simulations in the regions of the greatest slope and combining the additional probability distributions thus obtained with the probability distributions from the initial SMD simulation in Eq. (7). An advantage to be found in the application of WHAM to SMD data is that it has been shown that the WHAM equations (7), (8) generalize in a straightforward fashion to multiple dimensions [18].

The application of the van Kampen Ω-expansion to SMD analysis, in the form of the Gaussian drift method, has been found to be valid. When the friction coefficient was available, the method performed well; however, the requirement of careful use of smoothing makes it less robust than other methods we examined. The use of fluctuation information

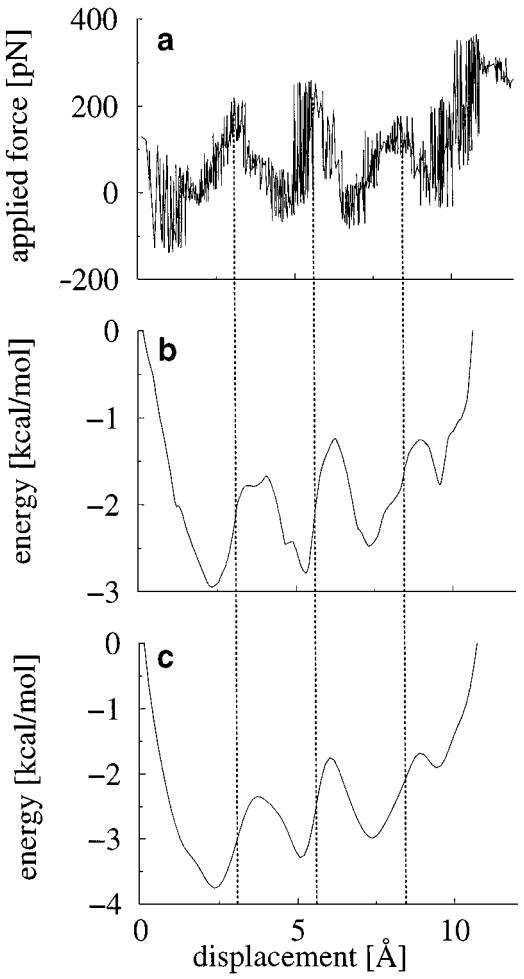

FIG. 10. Gaussian drift method and action minimization method applied to phospholipid membrane monolayer system (see [6]) data from an SMD simulation. (a) Smoothed force; (b) PMF reconstructed using the Gaussian drift method; (c) PMF reconstructed using the full action minimization method with clamped friction coefficient $\gamma = 4000$ pN · ps/Å.

was unfortunately of limited utility for the potentials we examined. Fluctuation measurements might be of some utility in quantifying the error in the predicted PMF, though this question has not been pursued.

The action minimization methods, in either the least squares or full action implementations, seem to hold the most promise, primarily because of their ability to assimilate many forms of information about the system to aid in the reconstruction of the potential. The reconstructions using the least squares method with clamped basis functions demonstrated that knowledge of the total change in the PMF leads to a very accurate reconstruction of the potential, not only at the endpoints but around midrange features as well. The total change in PMF is quite often available, through free-energy perturbation techniques and/or experimental measurements. Action minimization methods could also be made more accurate by providing information about the total irreversible work done during the SMD simulation, either by fixing the friction coefficient, rather than treating it as an adjustable parameter, or

by measuring the increase in energy and change in temperature of the ligand-protein system. The friction coefficient could be obtained by comparing simulations done at different pulling rates, as in [11], or by analysis of the ligand's velocity autocorrelation function, as demonstrated here for a model system.

Another means by which information about the potential might be incorporated into the reconstruction is through the choice of basis functions. The reconstructions exhibited in this paper were based on a model of $-dU/dx$ that was linear in its coefficients c; if one is willing to give up this property and resort to more sophisticated minimization techniques, one can employ basis functions such as Gaussians which are nonlinear in the coefficients. If obvious features are visible in the time series $x(t)$ and $F(t)$, then a good guess can be made for, e.g., the locations of the Gaussians, and the minimization may lead to new information about the height of the barriers in the PMF.

The least squares method is presently implemented for Langevin models in the strong friction limit; the corresponding Onsager–Machlup action integral of Eq. (21) contains only information about ligand positions. The strong friction limit is valid as long as sampling is not affected by processes faster than velocity relaxation times, typically a picosecond; thus it is valid for most of the ligand motion in SMD simulations. However, barrier crossing processes which exhibit fast jumps in $x(t)$ may correspond to ballistic motion, which might require a description in terms of Langevin models which do not assume the strong friction limit. Such jumps are, in fact, often visible in SMD simulations [4]. In this case the time series analysis needs to be based on an Onsager–Machlup action involving momentum as well as position, as described, for example, in [31].

In closing we would like to mention a complementary approach for constructing a PMF from SMD data that has been suggested by Evans and Ritchie [32]. In this approach a PMF was constructed by smoothing time series of position and ligand–protein interaction energy data monitored in the SMD simulation reported in [4]. Future investigations should combine this approach with the time series analyses suggested above to derive a PMF consistent with a broader range of SMD data.

ACKNOWLEDGMENTS

The authors thank S. Izrailev for kindly providing the DLPE simulation data and for fruitful discussions. This work was supported by the National Institutes of Health (NIH PHS 5 P41 RR05969), by the National Science Foundation (NSF BIR 94-23827 EQ), and by the Roy J. Carver Charitable Trust.

REFERENCES

1. V. T. Moy, E.-L. Florin, and H. E. Gaub, Adhesive forces between ligand and receptor measured by AFM, *Colloids Surfaces* **93**, 343 (1994).
2. E.-L. Florin, V. T. Moy, and H. E. Gaub, Adhesion force between individual ligand-receptor pairs, *Science* **264**, 415 (1994).
3. H. Grubmüller, B. Heymann, and P. Tavan, Ligand binding and molecular mechanics calculation of the streptavidin-biotin rupture force, *Science* **271**, 997 (1996).
4. S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, Molecular dynamics study of unbinding of the avidin-biotin complex, *Biophys. J.* **72**, 1568 (1997).
5. B. Isralewitz, S. Izrailev, and K. Schulten, Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations, *Biophys. J.* **73**, 2972 (1997).
6. S. Stepaniants, S. Izrailev, and K. Schulten, Extraction of lipids from phospholipid membranes by steered molecular dynamics, *J. Mol. Model.* **3**, 473 (1997).

7. H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation, *Biophys. J.* **75**, 662 (1998).
8. W. Nadler, A. Brünger, K. Schulten, and M. Karplus, Molecular and stochastic dynamics of proteins, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7933 (1987).
9. E. Barth and T. Schlick, Overcoming stability limitations in biomolecular dynamics. I. Combining force splitting via extrapolation with Langevin dynamics in LN, *J. Chem. Phys.* **109**, 1617 (1998).
10. D. Kosztin, S. Izrailev, and K. Schulten, Unbinding of retinoic acid from its receptor studied by steered molecular dynamics, *Biophys. J.* **76**, 188 (1999).
11. S.-J. Marrink, O. Berger, P. Tieleman, and F. Jähnig, Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations, *Biophys. J.* **74**, 931 (1998).
12. A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1351 (1999).
13. H. Lu and K. Schulten, Steered molecular dynamics simulations of force-induced protein domain unfolding, *PROTEINS Struc. Func. Genetics*, in press.
14. S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten, Steered molecular dynamics, in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, edited by P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, Lecture Notes in Computational Science and Engineering (Springer-Verlag, Berlin, 1998), Vol. 4, p. 36.
15. M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten, Reconstructing potential energy functions from simulated force-induced unbinding processes, *Biophys. J.* **73**, 1281 (1997).
16. C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach, *Phys. Rev. E* **56**, 5018 (1997).
17. C. Jarzynski, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.* **78**, 2690 (1997).
18. B. Roux, The calculation of the potential of mean force using computer simulations, *Comput. Phys. Comm.* **91**, 275 (1995).
19. K. Schulten, Z. Schulten, and A. Szabo, Dynamics of reactions involving diffusive barrier crossing, *J. Chem. Phys.* **74**, 4426 (1981).
20. J. Shillcock and U. Seifert, Escape from a metastable well under a time-ramped force, *Phys. Rev. E* **57**(6), 7301 (1998).
21. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kolman, and J. M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comp. Chem.* **13**, 1011 (1992).
22. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam/New York, 1992).
23. L. Onsager and S. Machlup, Fluctuations and irreversible processes, *Phys. Rev.* **91**, 1505 (1953).
24. Torrie and Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comput. Phys.* **23**, 187 (1977).
25. Radmer and Kollman, Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes, *J. Comput. Chem.* **18**(7), 902 (1997).
26. K. L. C. Hunt and J. Ross, Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic Lagrangian, *J. Chem. Phys.* **75**(2), 976 (1981).
27. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer-Verlag, New York, 1985).
28. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge Univ. Press, New York, 1992).
29. A. R. Leach, *Molecular Modelling, Principles and Applications* (Addison–Wesley/Longman, Essex, 1996).
30. W. F. van Gunsteren, J. C. Berendsen, and J. A. C. Rullmann, Stochastic dynamics for molecules with constraints, Brownian dynamics of *n*-alkanes, *Mol. Phys.* **44**, 69 (1981).
31. R. Olender and R. Elber, Calculation of classical trajectories with a very large time step: Formalism and numerical examples, *J. Chem. Phys.* **105**(20), 9299 (1996).
32. E. Evans and K. Ritchie, Dynamic strength of molecular adhesion bonds, *Biophys. J.* **72**(4), 1541 (1997).