Molecular simulation of the reversible mechanical unfolding of proteins

Nitin Rathore, Qiliang Yan, and Juan J. de Pablo[a]
Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706

In this work we have combined a Wang–Landau sampling scheme [F. Wang and D. Landau, Phys. Rev. Lett. **86**, 2050 (2001)] with an expanded ensemble formalism to yield a simple and powerful method for computing potentials of mean force. The new method is implemented to investigate the mechanical deformation of proteins. Comparisons are made with analytical results for simple model systems such as harmonic springs and Rouse chains. The method is then illustrated on a model 15-residue alanine molecule in an implicit solvent. Results for mechanical unfolding of this oligopeptide are compared to those of steered molecular dynamics calculations. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1649314]

INTRODUCTION

Single-molecule experiments are becoming increasingly important in biophysics. Experimental techniques such as atomic force microscopy (AFM) and optical tweezers have provided valuable insights into the behavior of biomolecules such as proteins or DNA under the influence of mechanical stresses. Single protein AFM[1] has been successfully applied to study the mechanical properties of tandem modular proteins, such as titin. Optical tweezers have been used to study proteins[2] and to understand DNA mechanics[3] involved in cellular processes of replication and transcription. Complementary computational approaches [steered molecular dynamics (SMD)[4] and Monte Carlo techniques[5,6]] have also been used to interpret experimental observations with atomic-level resolution.

Steered molecular dynamics techniques have been particularly useful in the study of a wide variety of protein systems.[7–10] These calculations aim to reproduce AFM experiments by subjecting a molecule to a time-varying external force. That force is applied by pulling one end of the molecule with a cantilever spring. A central quantity of interest in these SMD calculations is the potential of mean force (PMF), or reversible work required to pull the molecule. These calculations strive to achieve true reversibility. Jarzynski[11] has shown that free energy differences can in principle be computed from the exponential averages of the irreversible work; fast pulling rates and the fluctuations associated with the cantilever spring, however, can limit the accuracy of SMD calculations. Also note that SMD methods are not suitable for study of discrete or lattice models of proteins.

In this work, a random walk technique is proposed to construct the potential of mean force associated with the mechanical unfolding of peptide molecules. The method is based on the Wang–Landau scheme[12,13] and permits fast and self-consistent calculation of the density of states or free energy associated with arbitrary processes.

Our calculations of the density of states are conducted in the framework of an expanded ensemble. We refer to the formalism as expanded ensemble density of states (EXEDOS) method. For the particular case of mechanical stretching of proteins, this expanded ensemble is expressed in terms of the end-to-end distance of the peptide molecule being stretched. We begin this work by establishing a correspondence between EXEDOS and SMD simulations for the simple case of a Rouse chain. Free energy changes associated with the stretching of a Rouse chain are computed using these methods and compared to exact analytical results. Having established the validity of our proposed formalism, we extend the application of EXEDOS to a simple peptide system described within a united atom representation and the CHARMM19 force field.[14] A helical oligopeptide, namely 15-mer alanine in the presence of an implicit solvent, is used for our calculations. Results are presented in the form of force-extension curves and the corresponding potentials of mean force.

METHODS

Theory

The potential of mean force (Φ) associated with a specified generalized reaction coordinate, $\xi(\mathbf{r})$, measures the free energy change as a function of $\xi(\mathbf{r})$ (where $\mathbf{r}$ represents a set of Cartesian coordinates). This potential is related to the probability density of finding the system at a specific value ξ of the reaction coordinate $\xi(\mathbf{r})$,

$$P(\xi(\mathbf{r}) = \xi) \equiv C e^{-\beta \Phi(\xi)}, \tag{1}$$

where C is a normalization constant. The methods of umbrella sampling and weighted histogram techniques rely on estimating free energy changes from this probability density function. In umbrella sampling, the free energy barriers are eliminated by artificially modifying the potential function. This bias is later removed and the simulation data are re-

[a]Author to whom correspondence should be addressed. Electronic mail: depablo@engr.wisc.edu

     5781

weighed to arrive at the correct probability distribution. The potential of mean force is then calculated from

$$\Phi(\xi) = -k_B T \ln P(\xi) + C. \tag{2}$$

Another class of methods arrive at the potential of mean force by calculating the derivative of the free energy with respect to the constrained generalized coordinate ξ in a series of simulation runs. A mean force, $\langle F \rangle_\xi = [-\partial(\Phi(\xi))/\partial(\xi)]$, can be integrated numerically to yield the corresponding PMF. These simulations involve constraints, and an appropriate correction term must therefore be accounted for in the calculation of the PMF. For a system with no constraints and consisting of N particles with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, the mean force[15] can be written as

$$\langle F \rangle_\xi = \frac{\left\langle \frac{\partial \mathbf{r}}{\partial \xi} \left[-\frac{\partial U}{\partial r} + k_B T \frac{\partial \ln |\mathbf{J}|}{\partial \mathbf{r}} \right] \delta(\hat{\xi}(\mathbf{r}) - \xi) \right\rangle}{\langle \delta(\hat{\xi}(\mathbf{r}) - \xi) \rangle}, \tag{3}$$

where U is the potential energy function and $\mathbf{J}$ is the Jacobian associated with the transformation from Cartesian to generalized coordinates. To compute the mean force acting on the end-to-end distance of a molecule, a suitable reaction coordinate is provided by $\xi = r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, where r_{ij} represents the distance between the two terminal sites i and j. The Jacobian, $|\mathbf{J}|$, is a function of the separation r_{ij} and can be taken out of the ensemble average to yield[16]

$$\langle F \rangle_\xi = \left\langle -\frac{\partial U}{\partial \xi} \right\rangle_\xi + \frac{2 k_B T}{\xi}. \tag{4}$$

The mean force therefore includes a contribution from the average mechanical force and another contribution arising from the variations of the volume element associated with the reaction coordinate ξ. The free energy change between two states ξ_1 and ξ_2 can be obtained by integrating Eq. (4), according to

$$\Phi(\xi_2) - \Phi(\xi_1) = \int_{\xi_1}^{\xi_2} \left\langle \frac{\partial U}{\partial \xi} \right\rangle d\xi - 2 k_B T \ln\left(\frac{\xi_2}{\xi_1}\right). \tag{5}$$

Constrained simulations rely on the calculation of the first term of Eq. (4) from a series of simulations conducted at different values of ξ. This average force is then corrected by adding the second term of Eq. (4) and then numerically integrated to give a PMF for the desired range of ξ.

The EXEDOS method proposed in this work offers several advantages over the conventional schemes outlined above. The probability distribution that dictates the walk in the ξ space is computed on the fly during a simulation in a self-consistent manner. The simulation is performed without any constraints, which means that the resulting weights can be used directly as in Eq. (2) to give the potential of mean force. One can also accumulate the forces acting on the particles that define the reaction coordinate and then use Eq. (5) to get the PMF. The computed PMF is therefore available as a continuous function of ξ.

Model

Harmonic spring and Rouse chain

The first part of this work aims to establish a correspondence of different simulation methods in the context of a simple harmonic spring and a Rouse chain. The harmonic potential function describing the spring is given by

$$U = k_s r^2, \tag{6}$$

where r is the elongation of the spring and k_s is a force constant. The probability distribution for the end-to-end vector $\mathbf{r}$ is given by[17]

$$P_{\text{eq}}(\mathbf{r}) = \left(\frac{k_s}{\pi k_B T}\right)^{3/2} e^{-(k_s/k_B T)(\mathbf{r} \cdot \mathbf{r})}, \tag{7}$$

and the mean square end-to-end distance is given by

$$\langle r^2 \rangle_{\text{eq}} = \frac{3 k_B T}{2 k_s}. \tag{8}$$

The Helmholtz free energy (A) can be computed as

$$A(r) = U - TS = k_s r^2 - 2 k_B T \ln r + C, \tag{9}$$

where C is a constant. For a Rouse chain comprising N sites (or $N-1$ springs) the corresponding expressions are[17]

$$P_{\text{eq}}(\mathbf{r}) = \left(\frac{k_s}{(N-1)\pi k_B T}\right)^{3/2} e^{-(k_s/k_B T)(\mathbf{r} \cdot \mathbf{r})} \tag{10}$$

and

$$\langle r^2 \rangle_{\text{eq}} = \frac{3(N-1) k_B T}{2 k_s}, \tag{11}$$

i.e., the Rouse chain behaves as a single harmonic spring with an effective force constant given by $k_s/(N-1)$. These two simple models are used to validate the EXEDOS formalism and analyze the results of steered MD calculations.

Protein model and implicit solvent

The EXEDOS method proposed in this work is used to calculate the PMF for an oligopeptide system. For these simulations we use the CHARMM19 force field with a united atom representation. We use the EEF1 model parameters,[18] where the partial charges on the amino acids are modified to neutralize the side chains and the patched molecular termini. The interactions between atoms are described by the following potential energy function:

$$U_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta)]$$

$$+ \sum_{\text{impropers}} K_\omega (\omega - \omega_{eq})^2$$

$$+ \sum_{\text{LJ}} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right]$$

$$+ \sum_{\text{Coulombic}} \frac{1}{4\pi\epsilon(r)\epsilon_0} \frac{q_i q_j}{r_{ij}} + U_{\text{solvation}}. \tag{12}$$

A 1–3 exclusion principle is used for the nonbonded energy. The 1–4 Coulomb interactions are scaled down by a factor of 0.4. This is consistent with the original parametrization of CHARMM19. A cutoff of 12 Å is used for both the electrostatic and van der Waals terms. A simple force shift scheme is employed for Coulombic interactions, and a simple cut and shifted potential is employed for Lennard-Jones interactions.

An implicit solvent model based on the solvent accessible surface area (SASA) is employed, with solvation parameters as proposed by Ferrara et al.[19] Electrostatic screening effects are approximated by a distance dependent dielectric function and a set of partial charges with neutralized side chains. The model assumes that the mean solvation energy is proportional to the SASA of the solute. For a solute having M atoms with Cartesian coordinates $\mathbf{r}$, the solvation term is given by

$$U_{\text{solvation}} = \sum_{i=1}^{M} \sigma_i A_i(\mathbf{r}), \tag{13}$$

where σ_i and $A_i(\mathbf{r})$ are the atomic solvation parameter and SASA of atom i, respectively. The computations of atomic solvation parameter and the SASA are performed as indicated in Ref. 19. The SASA model, however, only accounts for the free energy cost of burying a charged residue in the interior of a protein. The solvent screening effect is approximated by using a distance-dependent dielectric function, $\epsilon(r) = r$. While this is an oversimplified way of accounting for solvent polarization effects, it is consistent with the formulation of SASA model parameters and previous simulations of proteins. Given that the purpose of this work is to examine the performance of a new algorithm, we restrict our calculations to a model and system that have been thoroughly characterized in the literature.

More specifically, we consider a polyalanine molecule of 15 residues. At room temperature this oligopeptide adopts a stable α-helical conformation in an implicit solvent. By applying an external stretching force, the molecule can be forced to undergo a helix–coil transition.

Simulation methods

Steered molecular dynamics (SMD)

Steered MD simulations[4] mimic a single molecule AFM experiment, where one end of the molecule is fixed and the other end is pulled using a cantilever spring. A time dependent external force is applied by moving the other end of the cantilever spring at a constant velocity in the desired direction, which is usually taken as the vector joining the fixed atom to the pulled atom. Figure 1 shows the typical setup of a SMD simulation. The stretching in these computations is assumed to be unidirectional (say along $\hat{\xi}$), and the scalar force acting on the molecule of interest along that direction at any time t is given by

$$F = k_c(vt - (\xi - \xi_o)), \tag{14}$$

where k_c is the spring constant of the cantilever, $(\xi - \xi_0)$ is the displacement of the pulled atom from its original position, and v is the constant pulling velocity. This force is
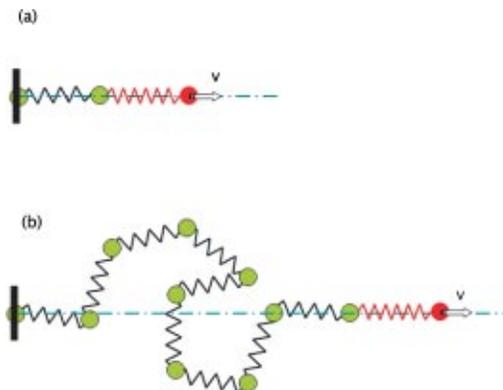

FIG. 1. Schematic representation of (a) a Hookean spring and (b) a Rouse chain being pulled in a steered MD setup. ξ is the end-to-end distance along the pulling direction denoted by vector $\hat{\xi}$.

collected as a function of the reaction coordinate, ξ, and can be integrated to yield the work done on the molecule. Assuming that the pulling speed is sufficiently slow, multiple SMD trajectories can in principle be averaged using Jarzynski's equality[11,20] to obtain the potential of mean force for the pulling process.

As discussed in the literature,[9,10] the spring constant of the cantilever can influence considerably the nature of the resulting force-extension curves. The system of interest and the cantilever spring must effectively behave as two springs in series. The spring constant of cantilevers typically employed in experiments is about 0.1 N/m.[21] In simulations, however, if the cantilever spring is too soft, it dominates the characteristics of the protein's response to a force. If the spring is too stiff, it leads to noise in the force estimation due to fluctuations of the position of the restraint point. There are no precise, à priori rules to determine the ideal stiffness constant for a specific system.

We propose to eliminate altogether the use of a cantilever spring from our calculations. The free end of the molecule need not be attached to a spring; it can be pulled directly at constant velocity along the direction of interest. It should be noted that pulling the molecule directly is not equivalent to using a very stiff spring. Eliminating the use of cantilever avoids the problems associated with stiff springs. Fluctuations in the reaction coordinate are reduced, thereby leading to a separation between the two ends that increases linearly with time. The work done on the system can now be calculated as

$$\Delta W(1 \rightarrow 2) = \int_{t_1}^{t_2} \mathbf{F} \cdot \mathbf{v} \, dt = v \int_{t_1}^{t_2} F_\xi \, dt, \tag{15}$$

where $\mathbf{F}$ is the total force acting on the particle being pulled and F_ξ is the component of this force along the pulling direction ($\hat{\xi}$). One can accumulate this component of the force during the course of a simulation and later multiply it by the pulling speed and time step to estimate the work.

A useful alternative proposed here to determine the work is to take advantage of the conserved nature of the Hamiltonian. A steered MD simulation can be performed with a

Nosé–Hoover thermostat.[22–24] For an NVT simulation using Nosé–Hoover chains, this conserved Hamiltonian ($\mathbf{H}'$) is given by[25]

$$\mathbf{H}'(\mathbf{p},\mathbf{q},\eta,\mathbf{p}_\eta) = H(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{M} \frac{P_{\eta_i}^2}{2Q_i} + N_f kT\eta_1 + \sum_{i=2}^{M} kT\eta_i, \tag{16}$$

where H is the regular Hamiltonian of the system, $\mathbf{q}$ and $\mathbf{p}$ represent the positions and momenta, η is a thermodynamic friction coefficient, and Q_i is the mass associated with a Nosé–Hoover chain, N_f is the number of degrees of freedom and M is the number of chains. In the absence of a pulling force this Hamiltonian ($\mathbf{H}'$) is conserved. If a SMD simulation is now performed by moving one end of the molecule at a constant velocity, the work done on the system can be measured directly from the rise in the Hamiltonian ($\Delta\mathbf{H}'$) for the deformation process. This eliminates the need for any post processing of simulation data. Also, the Nosé–Hoover Hamiltonian evaluated at every time step provides a more accurate estimate of the PMF than that determined by more conventional SMD approaches, which compute the work by integrating time averaged force-extension profiles. An added benefit of our proposed approach is that Nosé–Hoover thermostats give rise to more realistic dynamics than other temperature control schemes employed in the literature in the context of SMD simulations (such as simple velocity rescaling or Berendsen thermostats). It should be noted that the number of degrees of freedom of the system is now ($3N - 6$) due to the fact that the two ends are constrained to an axis.

Expanded ensemble density of states (EXEDOS)

In recent years, a new class of algorithms based on energy random walks[12,13] has emerged with the potential of providing a direct estimate of the density of states of a system, $g(E)$. Like multicanonical or umbrella sampling methods, these techniques are designed to eliminate free energy barriers, thereby permitting uniform sampling of phase space. Multicanonical methods often require nontrivial iterative calculations of weight factors; for umbrella potential methods to be effective, a good initial guess of the biasing potential is required. Density of states (DOS) methods circumvent these problems by computing a running best estimate of the biasing function. This function, $g(E)$, need not be known *à priori*; it is determined "on the fly" in a self-consistent manner. This scheme has been successfully applied to a variety of systems, including Ising models,[13] Lennard-Jones fluids,[26] proteins,[27,28] and glass-forming liquids.[29]

In recent work[30] we have explored the use of DOS methods in the context of expanded ensembles, where intermediate states are introduced to facilitate the transition between configurations separated by large energy barriers. The expanded states are usually defined by some reaction coordinate, ξ, and the sampling in ξ space is governed by unknown weights. This so-called EXEDOS method has been employed
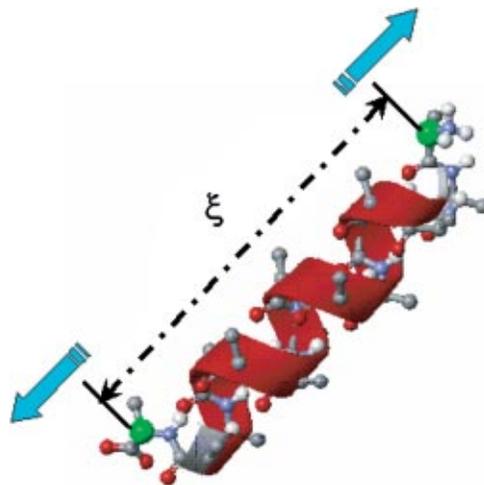

FIG. 2. Schematic representation of the oligopeptide, 15-residue alanine, being mechanically stretched in an EXEDOS setup. ξ is the end-to-end distance between the N and C terminus of the peptide molecule.

for studies of suspensions of colloidal particles in liquid crystals.[30] Here we use it to examine the reversible mechanical stretching of proteins. The reaction coordinate, ξ, is chosen to be the end-to-end distance (see Fig. 2) between the N and C terminus of the molecule being stretched.

The goal of the method is to perform a random walk in ξ space. Consider a system consisting of N particles interconnected to form a molecule, and having volume V and temperature T. The end-to-end distance of the molecule (ξ) can be discretized into distinct states; each state is characterized by its end-to-end distance, ξ, in some specified range of interest $[\xi_L, \xi_U]$; ξ_L and ξ_U represent a lower and an upper bound, respectively. The partition function Ω of this expanded ensemble is given by

$$\Omega = \int Q(N,V,T,\xi)g(\xi)d\xi = \int Q_\xi g(\xi)d\xi, \tag{17}$$

where Q_ξ is the canonical partition function for that particular state ξ, and $g(\xi)$ is the corresponding weight factor. The probability of visiting a state having extension ξ can be written as

$$P(\xi) = \frac{Q_\xi g(\xi)}{\Omega}. \tag{18}$$

The free energy difference between any two states can therefore be calculated from the weight factors, and the population density can then be determined according to

$$\Phi(\xi_2) - \Phi(\xi_1) = -k_B T \ln \frac{Q_{\xi_2}}{Q_{\xi_1}}$$

$$= -k_B T \left[\ln \frac{g(\xi_1)}{g(\xi_2)} + \ln \frac{P(\xi_2)}{P(\xi_1)} \right]. \tag{19}$$

If each state is visited with equal probability, the second term on the right-hand side of Eq. (19) disappears and the PMF can be computed from

$$\Phi(\xi) = -k_B T \ln g(\xi) + C. \tag{20}$$

J. Chem. Phys., Vol. 120, No. 12, 22 March 2004

Simulation of mechanical unfolding 5785

In EXEDOS, a running estimate of these weight factors can be computed and refined in a self-consistent manner. At the beginning of a simulation, $g(\xi)$ is assumed to be unity for all states. Trial Monte Carlo moves are accepted with probability

$$P_{\text{acc}}(\xi_1 \rightarrow \xi_2) = \min\left[1, \frac{g(\xi_1)}{g(\xi_2)}\exp(-\beta\Delta U)\right], \qquad (21)$$

where ξ_1 and ξ_2 are the end-to-end distances of the system before and after a trial move. After each trial move, the corresponding weight factor is updated by multiplying the current, existing value by a convergence factor f that is greater than unity ($f>1$), i.e., $g(\xi) \rightarrow g(\xi)\,f$. Every time that $g(\xi)$ is modified, a histogram $H(\xi)$ is also updated. This $g(\xi)$ refinement process continues until $H(\xi)$ becomes sufficiently flat. Once this condition is satisfied, the convergence factor is reduced by an arbitrary amount. Here we use $f_{\text{new}} = \sqrt{f_{\text{old}}}$. The histogram is then reset to zero ($H(\xi)=0$), and a new simulation cycle is started, continuing until the new histogram $H(\xi)$ is flat again. The process is repeated until f is smaller than some specified value, e.g., $f_{\text{final}}=\exp(10^{-8})$. Note that the weight factors, $g(\xi)$, change throughout the course of the simulation and detailed balance is not satisfied. Only towards the end of a calculation, when $f \rightarrow 1$, is detailed balance approached and the weight factors [or $g(\xi)$] converge to their "true," free-energy values. It must also be noted that, given that the convergence factor decreases as a calculation proceeds, configurations generated at different stages of the simulation do not contribute equally to the estimated density of states. In fact, towards the final stages of convergence, the convergence factor is so small that the corresponding configurations contribute only negligibly to the final estimate of $g(\xi)$.[31,32]

In addition to computing these weight factors from the histograms of visited states, one can obtain a second estimate from the integration of the mean force, as given by Eq. (5). The first term on the right-hand side of Eq. (4) can be estimated in an EXEDOS simulation. The component of the total force acting on the two sites that define the reaction coordinate along the end-to-end vector, $\hat{\xi}$, is accumulated as a function of ξ. At the end of the simulation this mean force is corrected by adding the corresponding second term of Eq. (4), and then integrated to yield the PMF.

As mentioned above, in the earlier stages of the simulation ($\ln f>10^{-5}$), when the convergence factor is large, detailed balance is severely violated. As a result, thermodynamic quantities computed during this time (including average forces) are incorrect. To avoid carrying this error into later stages, the accumulators for average forces are reset during the early stages. As the convergence factor decreases (e.g., $\ln f<10^{-5}$), the violation of detailed balance has a negligible effect, and the accumulators need not be reset anymore. All configurations sampled during the simulation now contribute equally to the force accumulator, thereby eliminating the problem of nonequal configurational contributions encountered in the original Wang–Landau scheme.

The Monte Carlo algorithm employed here comprises two types of trial moves. The first type consists of hybrid MD/MC displacements. At physiological temperature, these hybrid moves are not drastic enough to sample the ξ space efficiently and we therefore resort to a second type of Monte Carlo move. In this latter move, the molecule of interest is stretched along the end-to-end vector; the end-to-end distance is altered by some random amount and the coordinates of all the sites are rescaled accordingly.

To facilitate convergence and sampling, the ξ space can be fragmented into smaller overlapping domains or windows. Multiple, noninteracting replicas of the protein molecule can be created and simulated in different boxes. Each simulation box corresponds to a window with a specific range of ξ, and these end-to-end distance ranges are assigned so that windows corresponding to adjacent boxes overlap with each other. Configurations in different boxes are swapped at regular intervals during the simulation. This ensures that systems in individual windows do not get trapped in particular configurations as a result of the bounds imposed by the window size.

Once the simulation has converged and the correct weight factors have been estimated, one can run a production run starting with these converged $g(\xi)$. The system can therefore visit each state with equal probability and thermodynamic averages of properties relevant to one's system of interest can be sampled efficiently. In addition to computing average quantities as a function of extension, one can also sample any order parameter of interest. Instantaneous configurations collected during production runs can then be weighed by the corresponding weight factors and used to construct the energy landscape at a specific temperature.

RESULTS AND DISCUSSION

EXEDOS on a harmonic spring and a Rouse chain

We begin by discussing our results for the simple case of a harmonic spring. The free energy change that occurs when the spring is extended from ξ_1 to ξ_2 is given by

$$\Delta A(\xi_1 \rightarrow \xi_2) = k_s(\xi_2^2 - \xi_1^2) - 2k_BT\ln\left(\frac{\xi_2}{\xi_1}\right). \qquad (22)$$

A spring constant, k_s, of 418.4 kJ/mol/Å^2, and a temperature of 40 000 K were used in these calculations. A high value of T was chosen in order to exacerbate entropic contributions to the free energy [second term of Eq. (5)]. Figure 3 shows the potential of mean force, Φ, computed from the weights determined through EXEDOS and that estimated analytically using Eq. (22). Also shown is the estimate obtained by integrating the average force [with the correction term as in Eq. (5)]. The three estimates are in agreement with each other. These simple calculations help establish the validity of the EXEDOS method and emphasize the importance of the second term in Eq. (5). As we can see in Fig. 3 (dotted line), omitting this term can lead to significant deviation from the correct result.

Figure 4 shows analogous results for a Rouse chain comprising nine identical springs (each having a force constant of 418.4 kJ/mol/Å^2) at a temperature of 400 K. The PMF for a Rouse chain, which effectively behaves like a single spring with reduced effective spring constant, is calculated using
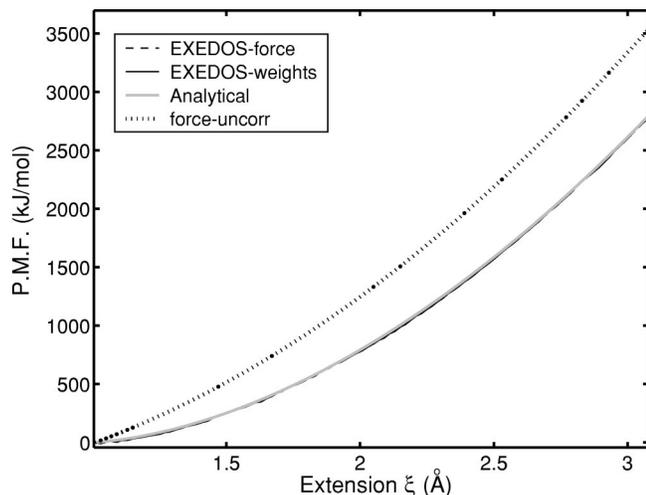
FIG. 3. Potential of mean force for a Hookean spring as derived analytically and from EXEDOS simulations.

FIG. 5. Comparison of discrepancies in potentials of mean force for Rouse chain obtained using different schemes.

the EXEDOS method, a conventional steered MD (SMD) technique (using a cantilever spring), and our proposed Nose–Hoover SMD (SMD–NH) variant, in which the last bead of the chain is constrained to move at a constant velocity along the pulling direction and the work is determined from the change in the Hamiltonian. We again emphasize that the results of SMD calculations must be corrected by the second term in Eq. (5). A cantilever spring with a force constant, k_c of 418.4 kJ/mol/Å^2 is used for SMD. The pulling velocities ($v = 0.0005$ Å/ps) used in our SMD and SMD–NH simulations are slow enough to approach the reversible pulling rate. The three estimates as shown in Fig. 4 are in good agreement with exact analytical results.

Useful insights are obtained into the two SMD schemes considered here. Figure 5 shows the deviation in the PMF estimates (SMD and SMD–NH) from the exact analytical value for two different pulling rates. The data from SMD–NH simulations exhibit considerably less statistical noise than those from conventional SMD simulations, primarily because the fluctuations associated with the cantilever
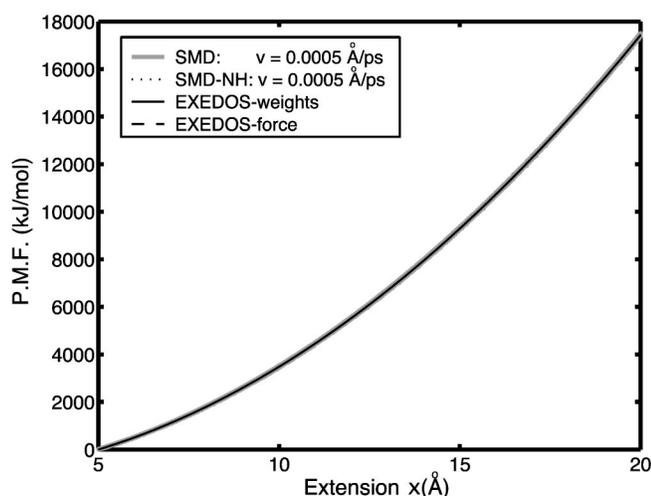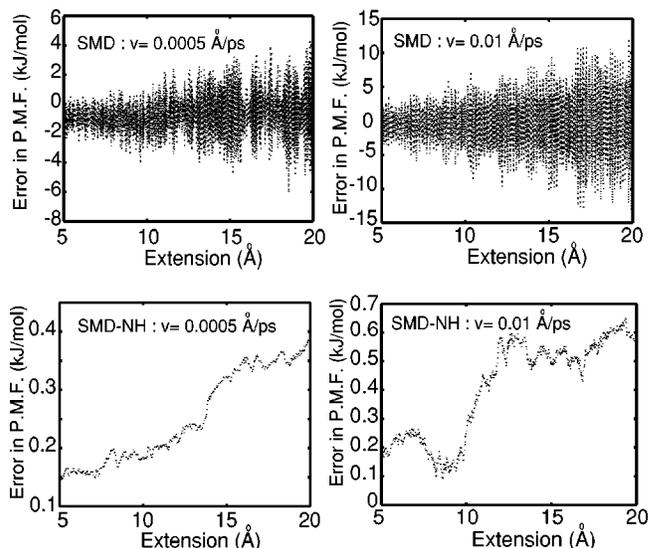
spring have been removed and the reversible work is calculated from the rise in the Hamiltonian. If Jarzynski's equality is employed to determine the PMF from multiple trajectories, these fluctuations are exponentiated and can result in large error bars. These findings suggest that the SMD–NH approach is better suited to be used in conjunction with Jarzynski's equality for calculation of free energies.

EXEDOS on proteins

Having established the validity of the formalism, in this section we consider its application to the study of reversible mechanical deformation of model proteins. The reaction coordinate, ξ, is chosen to be the end-to-end distance between the N and C termini of the peptide molecule being stretched. We present results for a 15-segment polyalanine molecule, which adopts a stable helical conformation in an implicit solvent.[19] The peptide is stretched at $T = 300$ K to a length much larger than that sampled in conventional canonical simulations (molecular dynamics or Monte Carlo) at this temperature. To facilitate convergence and sampling, the ξ space is again fragmented into smaller overlapping windows, each box representing a specific range of ξ that overlaps with that of adjacent boxes. Efficient sampling is achieved by using swap moves, in addition to regular MC moves. Figure 6 shows the PMF obtained for 15-mer alanine; the free energy minimum is arbitrarily set to zero. The ξ space explored in the simulation spans both the compression and tensile regimes. The forces therefore are negative when the helix is compressed to lengths smaller than the equilibrium end-to-end distance at 300 K. The PMF computed from the weights is in agreement with that computed by integrating the average force.

Figure 7 shows a comparison of free energy profiles obtained from SMD, SMD–NH, and EXEDOS calculations. The corresponding force-extension profiles for these calculations are shown in Fig. 8. The data presented for SMD and SMD–NH correspond to an individual trajectory. A cantile-

FIG. 4. Comparison of the potentials of mean force for a Rouse chain obtained using different schemes.
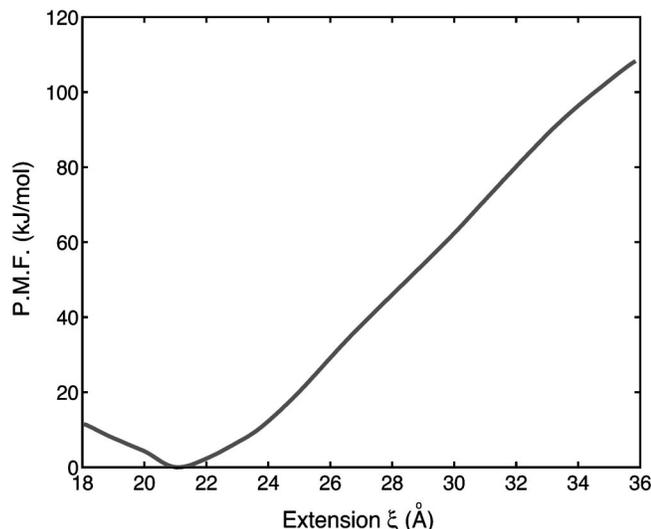
FIG. 6. Potential of mean force for 15-mer alanine as obtained from EXEDOS. The estimates computed directly from weights and by integrating the average force were in agreement with each other.

FIG. 8. Mean force-extension profiles for 15-mer alanine obtained using SMD and EXEDOS methods.

ver spring of spring constant, $k_c = 41.84$ kJ/mol/Å^2 (≈ 7 N/m) is used for SMD calculations. In order to arrive at a meaningful basis for comparisons between different schemes, the PMF for SMD and SMD–NH simulations is time averaged over 0.1 Å (the bin width in EXEDOS), thereby reducing some of the statistical noise. For higher pulling rates (rates comparable to those employed in the literature), the forces and the PMF obtained from steered MD without a cantilever (SMD–NH) exhibit less noise than those obtained from conventional SMD. As the pulling velocity is reduced, the results of SMD and SMD–NH approach the results of equilibrium, EXEDOS simulations.

In EXEDOS runs, once the simulation has converged and the correct weight factors have been estimated, one can run a production run starting with these converged weights. The system then visits each state with uniform probability

and relevant properties can be sampled efficiently. Figure 9 shows one such order parameter, the average helicity of the peptide molecule, as a function of extension. Stretching occurs by breaking the intrahelical hydrogen bonds. The average helicity decreases smoothly with extension rather than in discrete steps. The minimum in the PMF and the maximum in helicity is found at an extension of 21.2 Å. This corresponds to the helical conformation of polyalanine at room temperature in the absence of any constraint. As the molecule is pulled both the entropy and the potential energy increases. The rise in energy is faster than the gain in entropy, thereby resulting in an overall increase in the free energy of the molecule. Similar studies can be performed on small domains of protein molecules to obtain the potential along the unfolding coordinate. In addition, instantaneous samples can be collected during production runs and can then be weighed by the corresponding weight factors $[g(\xi)]$ to arrive at the energy landscape at any given temperature.
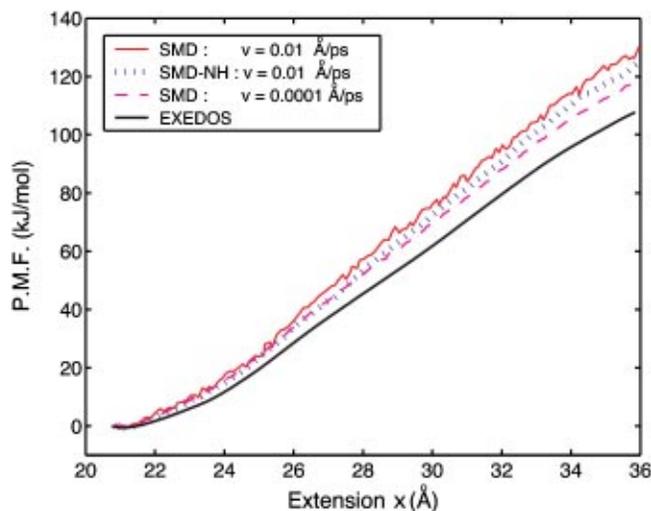

FIG. 7. Potentials of mean force as computed with SMD, SMD–NH, and EXEDOS simulations. As the pulling velocity is reduced, the SMD estimates converge towards the EXEDOS results.
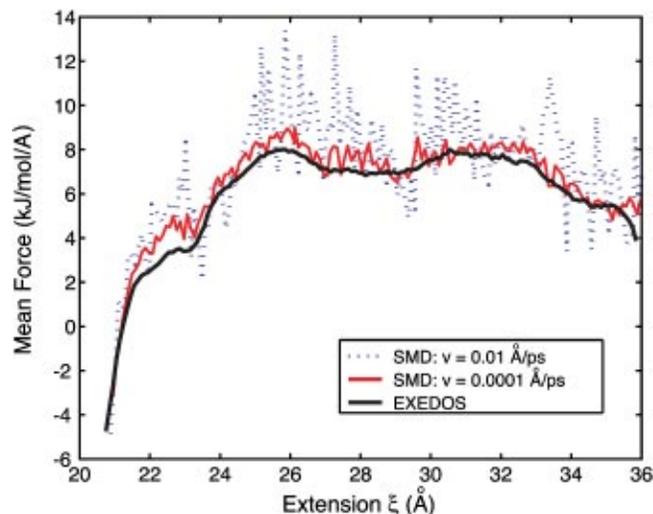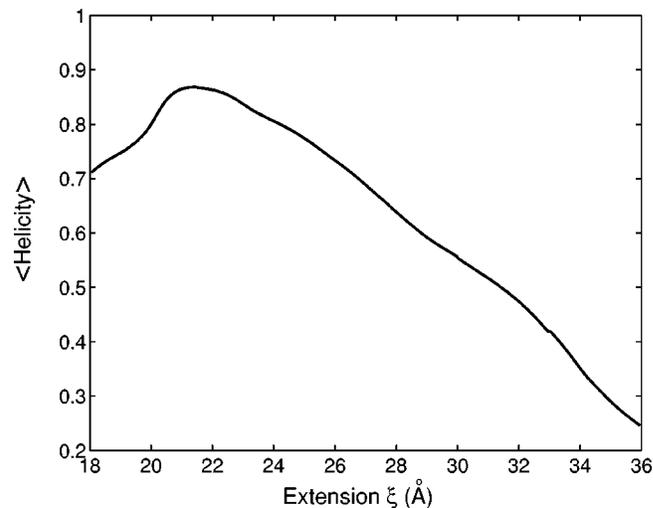

FIG. 9. Average helicity of 15-mer alanine as a function of extension at $T = 300$ K.

CONCLUSION

A new Monte Carlo method (EXEDOS) has been proposed for the study of the mechanical deformation of proteins. This method includes elements from both Wang–Landau sampling and expanded ensemble formalisms. Results have been presented for a model polyalanine molecule. In the presence of an implicit solvent, this molecule adopts a helical conformation. The proposed EXEDOS method is capable of stretching the helix to lengths much larger than those sampled in conventional canonical simulations, and provides reliable estimates of the potential of mean force required to deform the molecule. The uncertainty associated with EXEDOS simulations is considerably less than that encountered in steered molecular dynamics simulations. The EXEDOS approach also has the additional advantage of being equally applicable for systems in a continuum or on a lattice.

A simple variant of SMD has also been proposed that eliminates the use of a cantilever spring. Instead, in this variant (SMD–NH) the molecule is pulled by moving the terminal atom at a constant velocity. This reduces the noise associated with the fluctuations in the position of the restrain point, and permits direct calculation of the potential of mean force from the change in the Hamiltonian.

ACKNOWLEDGMENTS

[1] H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **97**, 6527 (2000).

[2] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, Science **276**, 1112 (1997).

[3] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature (London) **424**, 338 (2003).

[4] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, W. Wriggers, and K. Schulten, *Steered Molecular Dynamics in Computational Molecular Dynamics: Challenges, Methods, Ideas* (Springer-Verlag, Berlin, 1998), Vol. 4 of Lecture Notes in Computational Science and Engineering.

[5] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, UK, 1987).

[6] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kolman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).

[7] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662 (1998).

[8] B. Isralewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).

[9] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct., Funct., Genet. **49**, 104 (2002).

[10] Z. Bryant, V. S. Pande, and D. S. Rokhsar, Biophys. J. **78**, 584 (2000).

[11] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[12] F. Wang and D. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[13] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).

[14] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[15] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).

[16] E. Paci and M. Karplus, J. Mol. Biol. **288**, 441 (1999).

[17] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids* (Wiley, New York, 1987), Vol. 2.

[18] T. Lazaridis and M. Karplus, Proteins **35**, 133 (1999).

[19] P. Ferrara, J. Apostolakiz, and A. Caflisch, Proteins **46**, 24 (2002).

[20] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).

[21] H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez, Nat. Struct. Biol. **7**, 1117 (2000).

[22] S. Nosé, J. Chem. Phys. **81**, 511 (1984).

[23] S. Nosé, Mol. Phys. **52**, 255 (1984).

[24] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[25] G. J. Martyna, M. L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).

[26] Q. Yan, R. Faller, and J. J. de Pablo, J. Chem. Phys. **116**, 8745 (2002).

[27] N. Rathore and J. J. de Pablo, J. Chem. Phys. **116**, 7225 (2002).

[28] N. Rathore, T. A. Knotts, and J. J. de Pablo, J. Chem. Phys. **118**, 4285 (2003).

[29] R. Faller and J. J. de Pablo, J. Chem. Phys. **119**, 4405 (2003).

[30] E. B. Kim, R. Faller, Q. Yan, N. L. Abbott, and J. J. de Pablo, J. Chem. Phys. **117**, 7781 (2002).

[31] Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003).

[32] N. Rathore, T. A. Knotts, and J. J. de Pablo, Biophys. J. **85**, 3963 (2003).