Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · DECEMBER 2009

Impact Factor: 3.3 · DOI: 10.1021/jp9074898 · Source: PubMed

3 AUTHORS, INCLUDING:

Emanuele Paci
University of Leeds
124 PUBLICATIONS **6,453** CITATIONS
SEE PROFILE

Dmitry Shalashilin
University of Leeds
53 PUBLICATIONS **1,044** CITATIONS
SEE PROFILE

Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems

David R. Glowacki,*,†,‡ Emanuele Paci,† and Dmitrii V. Shalashilin‡

Centre for Computational Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom, and School of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom

Received: August 4, 2009; Revised Manuscript Received: October 8, 2009

We present a simple, comprehensive technique for accelerating simulation of rare events and calculating free energy profiles in molecular dynamics (MD) simulations. The technique is based on two related and complementary methods (AXD and BXD), which provide both thermodynamic and kinetic information along some reaction coordinate. The idea is to slice the reaction coordinate into several "boxes", and then run trajectories, locking them consecutively within each box to generate kinetic rate coefficients for exchange between neighboring boxes. In this way, a reaction coordinate may be efficiently explored, including those regions that would otherwise be visited only rarely. Combined with a simple and exact method for renormalizing the statistics obtained within each box, the box-to-box rate coefficients permit efficient free energy mapping. Tests carried out on model peptides demonstrate the utility of the method as well as the validity of the detailed balance assumption that underpins the derivations.

1. Introduction

The simulation of large molecules is a complicated field subject to numerous difficulties, but it continues to play an important role in biochemical and materials science. Despite many problems, molecular dynamics (MD) has proven a successful tool in simulating the behavior of large molecular systems. However, a significant limitation pertains to the fact that MD simulation of large molecular systems is limited to relatively short time scales, whereas events corresponding to chemical or large conformation change are rare.[1] A similar problem arises in calculating free energy profiles along an arbitrary reaction coordinate ρ. The normalized probability distribution of ρ, $p(\rho)$, is related to the projection of the free energy (or potential of mean force)[2] on ρ, $G(\rho)$, as follows:

$$p(\rho) = \frac{1}{\int \exp(-G(\rho)/kT)\, d\rho} \exp(-G(\rho)/kT) \qquad (1)$$

In principle, $p(\rho)$ may be calculated by running long time MD and computing the histogram of $\rho(t)$. However, accurately converging $p(\rho)$ in eq 1 for realistic systems is difficult in practice when regions of low probability (high free energy) are present and/or the diffusive dynamics of ρ is slow.

Although several methods have been proposed to overcome the sampling problem,[3−26] few are as straightforward, efficient, and easy to implement as the one we present herein. Among the reasons for this method's simplicity is the fact that no bias is added to the potential energy function of system, rendering a complicated unbiasing procedure unnecessary. Free energies and rate coefficients are among the most important calculable quantities to be obtained from MD; however, unlike the acceleration method presented in this article, few others simultaneously provide *both* kinetic *and* thermodynamic information. Furthermore, so long as the same assumption that underpins transition state theory (TST) is valid − that dynamics are randomized (or ergodic) on reaction time scales − the method described in this article is formally exact.

In what follows, we further develop the method of classical dynamics accelerated by phase space constrains[27,28] (henceforth, simply "accelerated dynamics" or AXD), and apply it to the simulation of biomolecules. In the AXD method, phase space[29] constraints are introduced along some reaction coordinate, dividing it into two "boxes", and the dynamics are locked within the box containing the transition state of interest. This speeds up the simulation and produces an accelerated rate of reaction that requires multiplication by correction factor.

In principle, the correction factor is just as costly to obtain as the corrected rate coefficient. Boxed molecular dynamics (BXD) is an efficient and seamless means for obtaining the correction factor and mapping free energy using a single long trajectory: it is an extension of AXD that partitions the reaction coordinate into more than two boxes, and shares similarities with milestoning.[15] Trajectories consecutively confined within each box generate kinetic rate coefficients for exchange between neighboring boxes. In BXD, the phase space along the reaction coordinate is systematically explored until the dynamics within each box are sufficiently well investigated. The global $p(\rho)$ is obtained using a simple formula to reweight the $p(\rho)$ obtained within each box, thus providing an accurate estimation of the free energy along the reaction coordinate, including those regions that would otherwise be visited only rarely.

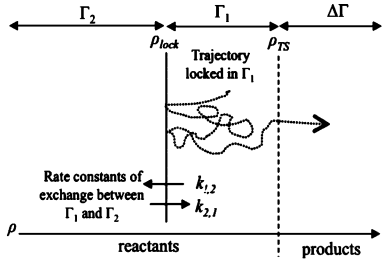
In the following section we provide a theoretical development of the techniques: (1) we reformulate our previous derivation of standard AXD, showing how it can be used to accelerate rare events; (2) we describe how to correct the accelerated AXD rate coefficients; (3) we describe the BXD technique; and (4) we describe a simple statistical renormalization procedure to be used with BXD.

Following the theoretical development, Section 3 demonstrates the utility of both AXD and BXD through simulations of loop formation for some model peptides. First, we describe implementation details and the model peptides. Second, we used BXD to

* Corresponding Author: drglowacki@yahoo.com.
† University of Bristol.
‡ University of Leeds.

10.1021/jp9074898
Published on Web 12/04/2009

SCHEME 1: Illustration of the AXD Approach for Calculating Accelerated Reaction Rates



calculate probability distributions and free energy profiles and compare computational efforts with those of brute force MD showing significant improvement of efficiency. Third, we perform AXD simulations to obtain rates of loop formation and use these simulations as a test of ergodicity and detailed balance between the boxes, which is the central assumption of both AXD and BXD. In the final section of the paper, we discuss the box boundary location, the significance of ergodicity to AXD and BXD accuracy, connections and comparisons with other existing methods of MD acceleration, and the system dependence of the efficiency gains.

2. Theory

2.1. Rare Event Acceleration with AXD. Classical dynamics accelerated with phase space constraints[27,28] (AXD) relies on the assumption that the classical mechanics of anharmonic systems is essentially randomized, or ergodic. So long as this is the case, then statistical mechanics is appropriate for calculating the probability of a rare event.

Scheme 1 illustrates the main ideas of AXD. Assuming transformation of reactants to products via some TS located at ρ_{TS}, we define $\Delta\Gamma$ as the phase space volume spanning the product side of the TS. In the AXD formulation, we separate the reactant phase space into two boxes: Γ_1, which spans ρ_{TS} to ρ_{lock}, and Γ_2, which is bounded by ρ_{lock}, the boundary between the two boxes (we use lowercase $p(\rho)$ to refer to the normalized probability of being at ρ, and uppercase $P(\rho)$ to refer to the integral of $p(\rho)$). The probability of being in $\Delta\Gamma$ may then be written as:

$$P(\rho \in \Delta\Gamma) = \int_{\rho\in\Delta\Gamma} p(\rho)\mathrm{d}\rho = \frac{\Delta\Gamma}{\Gamma_{tot}} = \frac{\Delta\Gamma}{\Gamma_1 + \Delta\Gamma} \times \frac{\Gamma_1 + \Delta\Gamma}{\Gamma_{tot}} = P^{AXD} \times P^{CORR}$$

$$P^{AXD} = \frac{\Delta\Gamma}{\Gamma_1 + \Delta\Gamma}$$

$$P^{CORR} = \frac{\Gamma_1 + \Delta\Gamma}{\Gamma_{tot}} \quad (2)$$

where Γ_{tot} is the total phase volume (i.e., $\Gamma_{tot} = \Gamma_1 + \Gamma_2 + \Delta\Gamma$), P^{AXD} is the enhanced probability of residing in $\Delta\Gamma$, and P^{CORR} is a correction factor. The fundamental efficiency gain of AXD derives from the claim that *it is easier to calculate P^{AXD} and P^{CORR} separately than their small product $P(\rho \in \Delta\Gamma)$*.

According to the classical theory of reaction rates, the phase space of the system is separated into reactant and product regions by a dividing surface at ρ_{TS} along some reaction coordinate, and the reaction rate coefficient is then calculated as a flux through the dividing surface. To obtain rate coefficients with AXD, the procedure is very similar to that in eq 2. The only difference is that $\Delta\Gamma$ for the correction factor effectively goes to zero because classical rate theory does not require a description of the phase space volume on the product side of ρ_{TS}. By locking the dynamics within Γ_1, transition state crossing events occur more often, yielding an accelerated rate of reaction, k^{AXD}. The actual rate coefficient, $k(T)$, may then be recovered as:

$$k(T) = k^{AXD} P^{CORR} \quad (3)$$

If the boxes are in detailed balance, then the rate coefficient correction factor is simply given by the ratio of the phase volume Γ_1 to the total phase volume (classical partition function) of reactants, $\Gamma_1 + \Gamma_2$, as follows:

$$P^{CORR} = \frac{\Gamma_1}{\Gamma_1 + \Gamma_2} = \frac{\int_{\rho\in\Gamma_1} p(\rho)\mathrm{d}\rho}{\int_{\rho\in\Gamma_1+\Gamma_2} p(\rho)\mathrm{d}\rho} \quad (4)$$

It can also be expressed through the equilibrium constant of exchange between the boxes, K_{12}, as follows:

$$P^{CORR} = \frac{1}{1 + K_{12}} \quad (5)$$

where

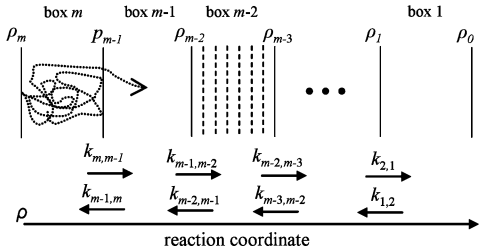
$$K_{12} = \frac{\Gamma_2}{\Gamma_1} = \frac{k_{12}}{k_{21}} \quad (6)$$

is related to the rates of "diffusion" between the boxes, and the rate constants k_{12} and k_{21} can be calculated from molecular dynamics within each box, as described below. Again, the rare event acceleration of these equations derives from the fact that *it is less expensive to converge k^{AXD} and P^{CORR} separately than their small product k(T) using brute force MD*. The Supporting Information contains further details of the derivation of eqs 2−6, as well as a discussion of how recrossing corrections may be included within the AXD formulation.

To confine trajectories within Γ_1, we have utilized a velocity inversion algorithm. At each integration time step t, we calculate the trajectory's position along the reaction coordinate. If it moves outside ρ_{lock} at time step $t + \mathrm{d}t$, then we return to the previous step t, and invert the components of the velocities in the direction of reaction coordinate for only those atoms that participate in the reaction coordinate definition. The inversion is carried out in a fashion that conserves the total energy, momentum, and angular momentum. Atoms whose velocities do not contribute to motion along ρ remain unchanged, and further details are provided in the Supporting Information.

2.2. Determining Free Energies with BXD. In this section, we present a generalization of AXD, which is called BXD (boxed molecular dynamics), and allows one to efficiently calculate P^{CORR} from restricted classical dynamics. In addition, it allows efficient and accurate mapping of the free energy, probability distribution, and kinetics along a particular reaction coordinate. In the BXD approach, the reaction coordinate is partitioned into a set of boxes, and restricted dynamics are simulated within each box just as in AXD. Trajectories are locked within each box using an inversion routine that preserves

SCHEME 2: Illustration of the BXD Procedure for a System Partitioned into *m* Boxes[a]



[a] In this simple picture, the trajectory penetrates from box m into box $m-1$ after two inversions at the ρ_{m-1} boundary, and the rate constant $k_{m,m-1} = h_{m-1}/t_m = 2/t_m$. In actual simulations, many more inversions are required to converge the rate constants. To illustrate the renormalization procedure, we have shown smaller histogram bins within larger boxes using dashed lines.

the integrals of motion, identical to that described above. By counting the number of inversion events with respect to the lifetime in each box, we then obtain a set of rate coefficients that describe transitions between adjacent boxes. The ratios of the box-to-box rate constants provide a set of box-to-box equilibrium constants, thus yielding ΔG for adjacent boxes. The entire set of ΔGs then allow one to map the PMF along the reaction coordinate.

The BXD approach is illustrated in Scheme 2, which shows the reaction coordinate ρ split into m intervals by boundaries at $\rho_0, \rho_1, ..., \rho_{m-1}, \rho_m$. Velocity inversion is carried out at each of the boundaries, and we keep track of the total time t the trajectory spends in a particular box. We also count the number of inversions at each box boundary. Assuming that we begin in box m, which is bounded by ρ_m and ρ_{m-1}, then the rate coefficient for transfer from box m to box $m-1$ is

$$k_{m,m-1} = \frac{h_{m,m-1}}{t_m} \tag{7}$$

where t_m is the lifetime of the trajectory in box m, and $h_{m,m-1}$ is the number of hits (i.e., velocity inversions) at the wall ρ_{m-1}. As illustrated in Scheme 2, we may relax the wall constraint following residence in box m and allow the trajectory to penetrate into the box $m-1$, which is defined by the boundaries ρ_{m-2} and ρ_{m-1}. Again, we keep track of the total time t_{m-1} that the trajectory spends in box $m-1$, and we count the number of hits $h_{m-1,m-2}$ and $h_{m-1,m}$ at each of the box boundaries. The respective rate coefficients for transfer from box $m-1$ to box m, and for box $m-1$ to box $m-2$ are then defined as follows:

$$\begin{aligned} k_{m-1,m} &= \frac{h_{m-1,m}}{t_{m-1}} \\ k_{m-1,m-2} &= \frac{h_{m-1,m-2}}{t_{m-1}} \end{aligned} \tag{8}$$

Both $k_{m-1,m}$ and $k_{m-1,m-2}$ represent the rate coefficients for diffusion through the two different box borders. Following residence in box $m-1$, the trajectory penetrates into box $m-2$, we perform a similar analysis, allow the trajectory to penetrate to box $m-3$, and carry on until the trajectory has visited all of the available boxes.

When all the box-to-box forward and reverse rate coefficients have been determined, and so long as a temperature may be defined, equilibrium constants between the neighboring boxes n and $n-1$ may be obtained as follows:

$$K_{n-1,n} = \frac{k_{n-1,n}}{k_{n,n-1}} = \exp\left(-\frac{\Delta G_{n-1,n}}{k_B T}\right) \tag{9}$$

The free energy difference between each neighboring box, $\Delta G_{n-1,n}$, may then be found by rearranging eq 9. With respect to some arbitrary zero, the free energy of each of the boxes, ΔG_n, may then be determined together with p_n, the probability of residing in box n:

$$p_n = \frac{1}{\sum_n \exp(-\Delta G_n/kT)} \exp(-\Delta G_n/kT) \tag{10}$$

The set of p_n defines the averaged free energies with resolution equal to the box size and allows one to renormalize the converged statistics obtained within each box and obtain the normalized probability distribution along the entire reaction coordinate, $p(\rho)$, to an arbitrarily high resolution. Letting $p_n(\rho)$ denote the probability that the reaction coordinate has some value ρ within box n using an arbitrary histogram bin size (smaller histogram bins within a larger box are shown in Scheme 2 by dashed lines), and assuming that we have normalized the total probability within each box to unity, that is,

$$\sum_{\rho_{min}}^{\rho_{max}} p_n(\rho) = 1 \tag{11}$$

then $p(\rho)$ along the entire reaction coordinate—that is, spanning all of the boxes—may be simply recovered by multiplying $p_n(\rho)$ by the normalized probability of residing in box n:

$$p(\rho) = p_n(\rho) \times p_n \tag{12}$$

Having obtained a normalized $p(\rho)$ using BXD, it is straightforward to work out the AXD correction factor P^{CORR} defined in eq 3 for an arbitrary ρ_{lock} as follows:

$$P^{CORR} = \frac{\int_{\rho\in\Gamma_1} p(\rho)\mathrm{d}\rho}{\int_{\rho\in\Gamma_1+\Gamma_2} p(\rho)\mathrm{d}\rho} \tag{13}$$

The BXD procedure uses kinetic transitions between nearest neighbor boxes in order to determine both $p(\rho)$ and $G(\rho) = -k_B T \ln p(\rho)$, and convergence can be assessed by examining the box-to-box rate coefficients. The time that the trajectory spends in each box is thereby determined only by how long it takes for these rate coefficients to converge. A particularly attractive feature of BXD is that it not only provides thermodynamic information, but also preserves kinetic information along the reaction coordinate. When all the box-to-box rates of diffusion are obtained, the global time dependence along the reaction coordinate boxes may be described using a set of coupled first order differential equations.

$$\begin{aligned}
\frac{dn_1}{dt} &= -k_{12}n_1 + k_{21}n_2 \\
\frac{dn_2}{dt} &= k_{12}n_1 + k_{32}n_3 - k_{21}n_2 - k_{23}n_2 \\
&\cdots\cdots \\
\frac{dn_m}{dt} &= k_{m-1,m}n_{m-1} - k_{m,m-1}n_m
\end{aligned} \tag{14}$$

Equation 14 is a discretized kinetic master equation (ME),[30,31] which determines the time-evolution of the probability distribution along the reaction coordinate.[32]

Just like AXD, the BXD approach assumes equilibrium dynamics between the boxes. That is, the inversion procedure implicitly assumes that, for every box, the incoming trajectories are compensated by the outgoing trajectories, and for every trajectory exiting the box there is a similar trajectory crossing the border and entering the box. By inverting the trajectory on the box boundary, we invert its velocity in the direction of the reaction coordinate without perturbing the velocity in other directions. This procedure is equivalent to jumping off the exiting trajectory and onto the incoming one, which does not disturb the equilibrium between the neighboring boxes. The dynamics acceleration achieved using BXD strongly depends on the system. For systems with large free energy barriers and/or slow diffusion, the BXD efficiency gain will be substantial, because the system will sample improbable regions as often as regions with a high residence probability.

In what follows, the AXD and BXD techniques are demonstrated through simulation of peptide cyclization, for which the reaction coordinate is the peptide extension. This simple coordinate provides us a straightforward test-bed for presenting these techniques; however, it is important to note that many other coordinates are possible—for example, gyration radius, number of native contacts, potential energy, or the distance between two molecular groups. AXD and BXD may also be generalized to phase space reaction coordinates whose description requires not only atomic positions, but also momenta—for example, the total energy of a particular bond or molecular fragment. In cases when a meaningful reaction coordinate is hard to define, ρ may simply be a progress parameter that parametrizes the process under consideration.

Finally, using a very similar development to that discussed above, the extension of AXD and BXD to multidimensional reaction coordinates for free energy mapping should be straightforward. For example, in the case of a two-dimensional reaction coordinate, each box will have a maximum of four boundaries (two in each dimension). However, the propagation of the dynamics through the boxes may still be carried out in a very similar manner to the "oscillating" implementation described below. Because the equations for deriving box averaged probabilities and carrying out renormalization do not depend on the precise definition of the reaction coordinate, they should be directly applicable to the multidimensional case.

3. Computational Details and Results

3.1. Implementation. In this section, we illustrate the AXD and BXD methods. The progress variable ρ on which we focus is peptide extension—that is, the distance between the main-chain nitrogen on the N-terminal residue and the carboxyl carbon of the C terminal one, which have been investigated previously using brute force MD.[33,34] The BXD implementation that we have used is entirely seamless insofar as we can use a single long trajectory to obtain the box-to-box rate coefficients without the need for reinitialization within each box. For each trajectory, we begin the simulation in a particular box. After some specified number of inversion events occur within that box, we turn off the velocity inversion and allow the trajectory to penetrate through the boundary that separates the first box from the next box, as shown in Figure 1. Following passage through the boundary, the velocity inversion routine is switched on again. When the trajectory reaches one of the two boxes that represent the reaction coordinate extrema, it then travels back through the boxes in the opposite direction from which it arrived. In this fashion, the trajectory traverses from one extrema of the reaction coordinate to the other in an oscillatory fashion, and we can monitor the convergence of the rate coefficients on each pass through the boxes until good convergence is achieved.

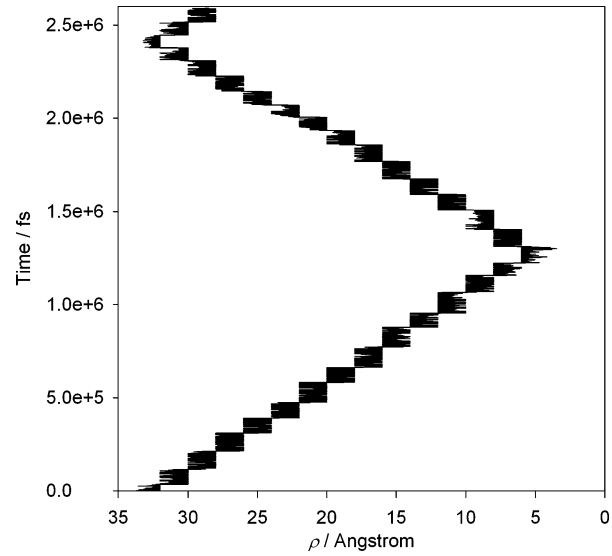


Figure 1. This figure illustrates oscillatory passage through the boxes by showing one complete oscillation—i.e., from large values of the extension to small values, and back again.

3.2. Systems. In this work, we present results for two different model peptides. The first is a 13 residue peptide with the sequence THR-TRP-ILE-GLN-ASN-GLY-SER-THR-LYS-TRP-TYR-GLN-ASN, and the CHARMM19[35] force-field with all attractive and electrostatic interactions switched off. Such a soft sphere model gives a relatively smooth free energy surface dominated by entropy. The second model is 10-alanine with the CHARMM19 force-field and a simple implicit solvation model.[36] With this force-field, 10-alanine turns out to have a strong helical propensity. The MD was run for 1 μs using Langevin dynamics with a friction coefficient of 1 ps^{-1} and a time step of 1 fs. In what follows, we refer to the 13 residue peptide as model 1, and the 10 residue alanine peptide as model 2. Simulations have been performed at 330 K for model 1 and 300 K for model 2.

3.3. Mapping Free Energies and Probabilities. Figure 2a shows BXD results for the box averaged free energy, G_n, and the integrated probability distribution, P_n, along the extension reaction coordinate, ρ, for model 1. The vertical lines in Figure 2a indicate where we located box boundaries. In the top panel of Figure 2a, the points within each set of boundaries correspond to the box averaged free energy in units of k_BT (T = 330 K) obtained by rearranging eq 9 to solve for $\Delta G_{n-1,n}$ and choosing an arbitrary zero. In the lower panel of Figure 2a, the points correspond to the box averaged P_n, obtained by integrating over the appropriate boxes for p_n in eq 10.

The resolution of Figure 2a is determined by the box boundaries; however, renormalizing with eq 12, we may obtain

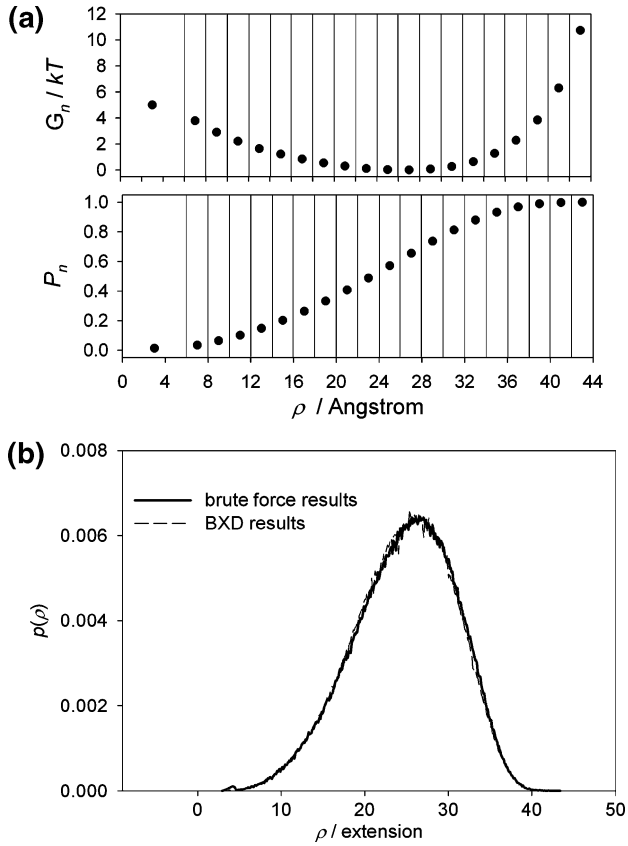


Figure 2. (a) Box averaged G_n and P_n for model 1. (b) The dashed line shows BXD results for $p(\rho)$ in model 1 renormalized using eq 12, which are essentially indistinguishable from the converged brute force results (solid line).

plots to arbitrary histogram resolution. Figure 2b shows $p(\rho)$ at a resolution of 0.1 Å. The small cutoff for calculation of nonbonded terms in the potential energy function yields a relatively smooth free energy surface for this peptide, so that the increased resolution plots shown in Figure 1b show little additional structure beyond the smooth structure in the courser box averaged plots. As shown in the figure, the BXD results renormalized using eq 12 are essentially indistinguishable from the converged brute force MD.

Figure 3a shows box averaged plots of both G_n and P_n for model 2. Unlike the peptide in Figure 1, it has a well-defined minimum at around 15 Å due to the high probability for this peptide to be in a helical conformation.

The plots in Figures 3b and 3c show $G(\rho)$, $p(\rho)$, and $P(\rho)$ renormalized using eq 12. The minima at ~4 and 6 Å correspond to van der Waals contact wells for peptide loop formation. The plots of $p(\rho)$ and $P(\rho)$ show a very low probability for both loop formation and for extension larger than ~20 Å.

Convergence criteria are well-defined within the oscillating implementation of BXD described above. With each pass through the boxed reaction coordinate, the free energy surface and probability distributions may be updated, and the simulations terminated upon arrival at some predefined convergence criteria. This is illustrated in Figure 4, which shows p_n, the box averaged probability that model 2 resides in a particular box, obtained using eq 10 after 1, 5, and 30 oscillations through the entire boxed reaction coordinate. The results with 30 oscillations are "fully converged" insofar the value of p_n changes by less than 1% with subsequent oscillations. Indeed, this level of convergence is obtained much earlier on, after only 18 oscillations.

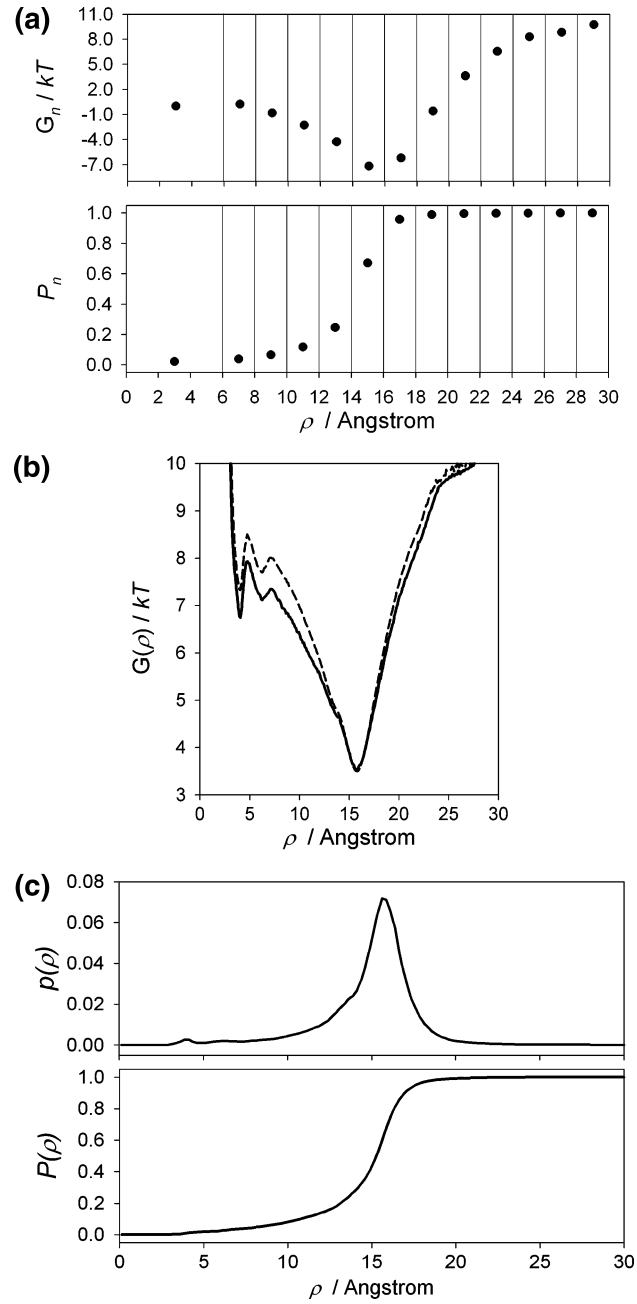


Figure 3. (a) Box average G_n and P_n for model 2. (b) Renormalized plot of $G(\rho)$ for model 2. The converged BXD results (solid line, 30 oscillations) are compared with brute force results for the same amount of CPU time (dashed line). As the figure shows, the brute force results do not converge at small and large extensions. (c) Renormalized plot of $p(\rho)$ and $P(\rho)$ for 10-alanine peptide.

Of course, the required level of convergence depends on the problem at hand, and Figure 4 shows that the results obtained after one oscillation are a reasonable approximation to the fully converged results.

The improved rare event sampling of BXD with respect to brute force simulation is shown in Figure 3b, which compares the converged high resolution plot of $p(\rho)$ for model 2 obtained after 30 BXD oscillations to brute force results obtained for the same amount of CPU time (~60 h). As expected, the brute force simulations converges poorly in the presence of high free energy barriers—namely, small and large values of the extension coordinate.

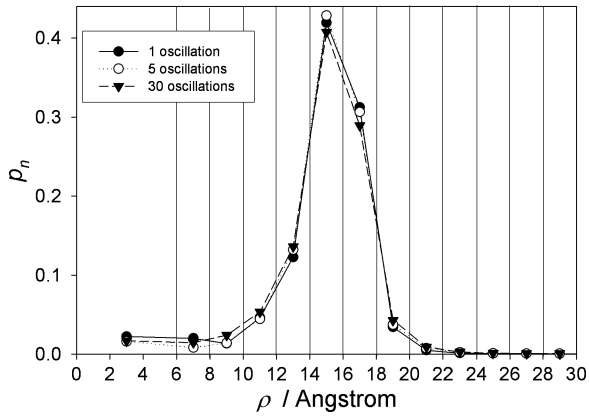


Figure 4. Plot of box averaged $p(\rho)$ for model 2, showing convergence with the number of oscillations through the boxed reaction coordinate.

3.4. AXD Rare Event Acceleration and Ergodicity Tests. Using AXD in conjunction with BXD, we performed simulations of loop formation kinetics. Table 1 gives rates of loop formation for both model 1 and model 2 obtained with AXD simulation using a range of values for ρ_{lock}. For model 1, we placed the TS at a separation of 5 Å, and for model 2, the TS dividing surface was located at 4.7 Å, near the free energy barrier to contact formation shown in Figure 3b. In addition to the accelerated rate k^{AXD}, Table 1 gives the values of P^{CORR} (obtained using eq 13) and the corrected rate coefficient $k^{\text{AXD}}P^{\text{CORR}}$ (also plotted in Figure 5). Within the two standard deviation error limits shown in the plot, one sees that the corrected rate coefficient for contact formation, $k^{\text{AXD}}P^{\text{CORR}}$, does not depend on ρ_{lock}.

This provides justification for the assumption of detailed balance and ergodic nature of the dynamics which is implicit in eq 4, 5, and 9. That is, the AXD results corrected using BXD correction factors show that the AXD and BXD boxes are in equilibrium.

Table 1 also illustrates the computational efficiency of AXD. The smaller the value of ρ_{lock}, the larger the number of reaction events. For example, in model 1, with $\rho_{\text{lock}} = 12$ Å then there are 3992 contact events, compared to 172 contact events with essentially unrestricted dynamics ($\rho_{\text{lock}} = 40$ Å). This simple comparison shows that obtaining a similar number of reaction events using unrestricted dynamics would require roughly 20 times longer CPU time.

An important feature of AXD is that it preserves certain nonstatistical effects[27] of the sort recently reviewed by Louderaj and Hase.[37] Within an equilibrium canonical ensemble of initial conformations, MD simulations sometimes show a fraction of very fast events, resulting in a nonexponential distribution of reaction times. Although we do not discuss such nonstatistical effects in detail herein, such events are preserved by AXD and will be addressed further in a future article.

4. Discussion

4.1. Location of Boundaries. There are certain limitations on how small a constraining box can be. Assuming that the dynamics of large molecular systems is mainly stochastic except for the short time dynamics at the time scale below the correlation time τ_{corr}, then it follows that the lifetime in the box should exceed τ_{corr}. Restated, the size of the box, $\Delta\rho$, should be larger than the correlation distance

$$\Delta\rho \gg \langle\dot{\rho}\rangle\tau_{\text{corr}} \tag{16}$$

where $\langle\dot{\rho}\rangle$ is the thermally averaged velocity along the "reaction coordinate". The good agreement of our results with those of standard simulation suggests that they do not suffer significant perturbation on account of boundaries located too near one another.

So long as the box size is larger than that discussed above, then a natural question is where to locate the box boundaries. Because the AXD and BXD approaches are exact for ergodic systems, they represent nothing more than an efficient way of gathering statistics. Consequently, all of our tests thus far show that $\Delta G(\rho)$ and $p(\rho)$ are insensitive to the location of the box boundaries within statistical error limits. For the test systems examined herein, the approach that we have adopted is to simply choose an initial set of box boundaries. Of course, the efficiency with which a trajectory travels to a neighboring box depends on the underlying free energy landscape along the reaction coordinate. That is, if there is a large free energy gradient that decreases the likelihood of passage from one box to another, then it may be necessary to change the box width to allow the trajectory to penetrate the boundary more quickly and so arrive at the next box. Box boundary locations that yield significant efficiency gains may therefore involve some trial and error; however, in the simulations that we present here, location of box boundaries did not present a significant complication, and such effects are generally easy to spot because they are localized to the most rugged regions of the free energy landscape.

4.2. AXD/BXD, Ergodicity, and Other Approaches. Just as TST assumes that the dynamics of a particular molecular system is statistical, or ergodic, the same is true for the AXD and BXD techniques. For both AXD and BXD, the idea is to lock the dynamics in a certain area of the phase space and investigate the dynamics at short time scales. The

TABLE 1: AXD Results of the Loop Formation Rate for Both Models 1 and 2, Obtained with Different Values of ρ_{lock}[a]

model peptide	ρ_{lock} (Å)	No. of contact events	k^{AXD} (s^{-1})	P^{CORR}	$(k^{\text{AXD}}P^{\text{CORR}}) \pm \sigma$ (s^{-1})
model 1	12	3992	1.16×10^{10}	0.026	$(2.98 \pm 0.17) \times 10^{8}$
	16	1434	3.07×10^{9}	0.089	$(2.74 \pm 0.23) \times 10^{8}$
	24	322	6.64×10^{8}	0.437	$(2.90 \pm 0.72) \times 10^{8}$
	32	192	3.39×10^{8}	0.895	$(3.03 \pm 0.72) \times 10^{8}$
	40	175	2.25×10^{8}	0.999	$(2.25 \pm 0.87) \times 10^{8}$
model 2	10	13 054	2.87×10^{10}	0.021	$(1.29 \pm 0.65) \times 10^{9}$
	12	7653	8.77×10^{9}	0.045	$(1.05 \pm 0.74) \times 10^{9}$
	16	2340	1.50×10^{9}	0.099	$(1.05 \pm 0.50) \times 10^{9}$
	20	1328	7.93×10^{8}	0.649	$(7.90 \pm 3.60) \times 10^{8}$
	24	1308	1.12×10^{9}	0.985	$(1.12 \pm 0.36) \times 10^{9}$

[a] The corrected rate coefficients, $k^{\text{AXD}}P^{\text{CORR}}$, are given in the final column and plotted in Figure 5.

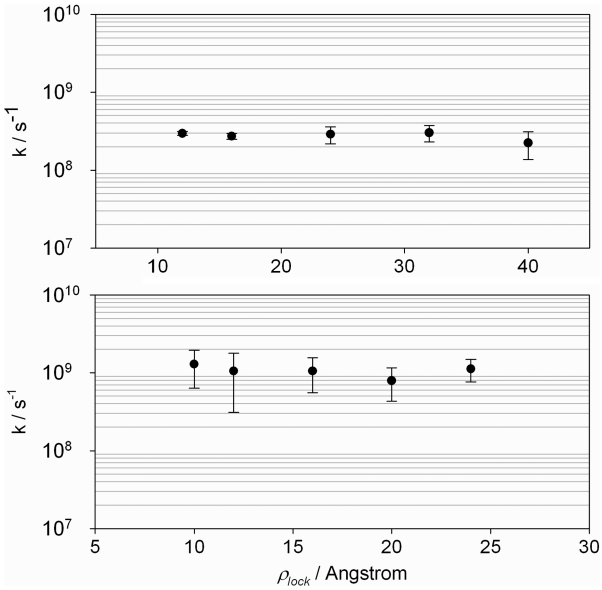


Figure 5. Corrected rate coefficients for contact formation in model 1 (top panel) and model 2 (bottom panel) at different values of ρ_{lock}. The corresponding data is given in the last column of Table 1.

difference between AXD and BXD is that the former is focused on calculating rates while the latter is concentrated on mapping the free energy, thereby providing an efficient means of obtaining P^{CORR}. BXD slices the phase space into boxes and systematically explores it box by box until the dynamics within each box is converged and a local free energy is obtained within the box. After that, the kinetics of exchange between the boxes is used to reconstruct the box averaged reaction coordinate free energy profile, and the high resolution statistics within each box are renormalized to obtain the global free energy. In this fashion, we explore regions rarely visited in direct simulation.

There are several methods reported in the literature for increasing the likelihood of rare events and mapping free energy landscapes.[5−21] As discussed elsewhere,[27] AXD is related to the method of hyperdynamics proposed by Voter,[11−13] which introduces boosting potentials to push the dynamics toward the product region just the way trajectory inversion does on the boundary between the boxes. In contrast to hyperdynamics, our formulation of the AXD method shows very clearly that it arises from the ergodicity of classical mechanics in big systems. AXD has some advantages with respect to hyperdynamics: (1) the boxes and reaction coordinates may be positioned in phase space (e.g., we may consider the reaction coordinate to be the energy of a molecular fragment),[27] whereas hyperdynamics boosting potentials are necessarily in coordinate space only; (2) AXD requires modification of neither the potential energy function nor the forces involved in the MD simulation.

BXD shares similarities with the methods of umbrella sampling,[6,14] and the z-constraint algorithm,[38] both of which provide thermodynamic information. In umbrella sampling, local dynamics are biased with harmonic potentials so that they evolve along selected portions of the reaction coordinate, and the free energy is recovered by means of the weighted histogram method (WHAM).[10] In the z-constraint algorithm, the dynamics are unrestricted in the x and y coordinates, but confined to discrete values along the z coordinate. The free energy at each z point is then recovered by calculation of the force. The introduction of velocity inversion in BXD and subsequent use of eq 12 in order to reconstruct the global probability distribution from a set of local statistics within each box is exact, and simpler than WHAM, directly exploiting the ergodicity of the system to provide not only thermodynamic information, but also kinetic information. In addition, as discussed in the previous paragraph, a unique feature of AXD and BXD is that they may be generalized to allow the use of phase space reaction coordinates—that is, the boundaries between the boxes may be a function of not only the atomic coordinates, but also of their momenta. For example, in this fashion, one could run AXD and BXD simulations and use the total energy (i.e., kinetic + potential energy) of a particular bond as a "reaction coordinate". This possibility stands in contrast to the harmonic biasing potentials used in umbrella sampling, which are functions only of the atomic coordinates.

Another method that shares similarities with our BXD technique is milestoning,[8,9,15] which uses short time dynamics to recover local kinetic information about the motion along the reaction coordinate in order to solve a kinetic state-to-state master equation. The original milestoning procedure is complicated because it requires accurate sampling of the equilibrium distribution at the hyperdimensional milestone surfaces located along the reaction coordinate.[9] In a recent paper,[8] a modified milestoning procedure has been proposed wherein the dynamics initiated between a set of milestones are then locked so that they cannot escape; however, the methods detailed in this paper differ from and develop some aspects of the modified milestoning procedure. First, we have demonstrated the link between BXD and AXD, showing how k^{AXD} may be corrected through P^{CORR}, independent of the location of ρ_{lock}. We have also shown how to simply and exactly renormalize the BXD statistics to obtain a free energy profile to arbitrary histogram resolution. Additionally, unlike the locking method used in that work, our velocity inversion technique preserves the linear and angular momentum of the system, so that it may be used for microcanonical ensembles. Finally, we have introduced a convenient and seamless sampling method, using a single long trajectory that oscillates through the reaction coordinate boxes, successively penetrating each box, and shown that it is not necessary to reinitialize sampling from individual points in the configuration space.

Previously, the idea of exploring short time dynamics in order to obtain kinetics of motion along the reaction coordinate has also been used in intramolecular dynamics diffusion theory (IDDT),[39−42] which describes the dynamics along the reaction coordinate in terms of a diffusional equation[41,42] or equivalent Langevin equation.[39] IDDT slices the reaction coordinate into boxes and determines the diffusional coefficient or equivalent Langevin friction coefficient from short time MD with initial conditions sampled box by box. Similar to IDDT, BXD and AXD produce the rates of diffusion to the neighboring boxes. The system of kinetic equations (eq 14) can easily be recast into a diffusional equation. In recent years a number of methods that run short time dynamics with initial conditions started at several regions along the reaction coordinate have appeared [e.g., partial path transition interface sampling (PPTIS), and forward flux sampling (FFS)]. In these methods, the short time dynamics results are recombined to recover rare event rate coefficients and/or free energies.[43−47] These methods share many similarities with IDDT, but unlike BXD, do not exploit the useful idea of restricting the dynamics itself—thus avoiding the need to reinitialize trajectories along the reaction coordinate.

A general property of any classical system is that its dynamics may contain ergodic and nonergodic contributions. The relationship between the two regimes and the origin of ergodicity is a

focus of a broad area of mathematics.[48,49] It is generally assumed that the ergodic regime dominates large molecular systems comprised of many strongly coupled degrees of freedom. Nonstatistical dynamics are often prevalent in small molecules,[37,50−52] and there is presently controversy regarding their significance in larger molecules.[53,54] Because the AXD and BXD methods rely upon an assumption of ergodic equilibrium dynamics, they offer a general indicator for when nonstatistical dynamics are important. We have obtained two results that indicate that the systems examined in this work are largely ergodic: (1) the BXD results are independent of the location of box boundaries within statistical error, and (2) the AXD rate coefficients corrected by the BXD probability distributions agree within statistical error, as shown in Figure 5 and Table 1.

4.3. Efficiency Gains. Simulating rates of slow reactions and obtaining distributions and free energy profiles like those discussed above using straightforward MD is often an expensive task depending on a number of system variables, for example, the treatment of solvent effects, the size and number of barriers along the reaction coordinate, and the simulation temperature. To obtain converged kinetic rate coefficients and mean first passage times, the simulation time required scales as the exponential of the barrier height. Because the efficiency gain obtained using AXD and BXD will vary significantly from system to system, it is difficult to provide a quantitative estimate of the general reduction in computational expense; however, for the relatively simple systems described above, the computational expense required to obtain converged results was reduced by approximately 1−2 orders of magnitude with respect to brute force MD, similar to the efficiency gains reported in the previous work using phase space constrained dynamics[27] for small gas phase molecules. The peptides chosen for simulation in this paper are sufficiently small for brute force MD to be feasible and available for comparison. For systems with larger free energy barriers, the speedup will be more significant.

Much of the MD literature has proposed solutions to the fact that time in MD is essentially a serial process and is not therefore easily parallelizable. For example, this is the problem addressed by methods such as IDDT or the distributed computing approach,[55−57] which attempt to extrapolate mean rates obtained from short time simulations to longer, experimentally accessible time scales. AXD in combination with BXD offer an approach for investigating this problem, as well as the accuracy of such approaches. BXD exploits the fact that free energy is a state function, and effectively simplifies the question of when to move along the reaction coordinate and into the next box according to some statistical convergence criteria. As such, BXD is easily parallelizable. Used in conjunction with the similarly easy-to-parallelize AXD, these methods provide an opportunity to efficiently simulate a range of systems and quantitatively investigate whether rates obtained from short-time dynamics are representative of long-time kinetics.

4.4. Conclusions. Free energies and rate coefficients are among the most important calculable quantities to be obtained from MD. We have presented a comprehensive approach to efficiently calculate reaction coordinate free energies, and rate coefficients for rare events. In the gas phase, these methods may enable the efficient calculation of free energy surfaces and location of free energy barriers without the need to calculate partition function approximations along the reaction coordinate, as required in variational transition state theory (VTST).[58,59] In addition, BXD may enable the efficient calculation of free energy surfaces for systems with poorly defined energy barriers, such as radical−radical addition reactions, which require a flexible transition state theory approach.[60] Used in conjunction with quantum dynamics approaches that utilize classical MD, BXD should even allow the calculation of free energy surfaces that fold in quantum effects.

The simulation of large biomolecules is a complicated field subject to numerous difficulties, not least of which concerns the choice of an appropriate reaction coordinate. However, the large choice of reaction coordinates which may be investigated using the approach detailed herein is clearly a great advantage. The simplicity of AXD and BXD makes their implementation in widely used biomolecular simulation MD packages (e.g., AMBER,[61] CHARMM,[62] GROMACS,[63] and NAMD[64]) unproblematic, facilitating its application to a large number of old and new problems.

Acknowledgment. We thank Zu Thur Yew, Emilio Martinez Nunez, Michael J. Pilling, Christopher Woods, Jeremy Harvey, Adrian Mulholland, Benoit Roux, Ron Elber, and Eric Vanden-Eijnden for useful discussions while preparing this article.

Supporting Information Available: Detailed derivation of the AXD formalism and details of the velocity inversion procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Ann. Rev. Phys. Chem.* **2002**, *53*, 291.
(2) In this paper, we refer to the *potential of mean force* and *reaction coordinate free energy profile* interchangeably.
(3) Zuckerman, D. M.; Woolf, T. B. *Phys. Rev. E* **2001**, *63*, 10.
(4) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904.
(5) Kinnear, B. S.; Jarrold, M. F.; Hansmann, U. H. E. *J. Mol. Graphics Modell.* **2004**, *22*, 397.
(6) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press: London, 2002.
(7) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120/1.
(8) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, 194101/1.
(9) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. *J. Chem. Phys.* **2008**, *129*, 174102/1.
(10) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(11) Voter, A. F. *J. Chem. Phys.* **1997**, *106*, 4665.
(12) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.
(13) Voter, A. F.; Montalenti, F.; Germann, T. C. *Ann. Rev. Mater. Res.* **2002**, *32*, 321.
(14) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(15) Faradjian, A. K.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880.
(16) Chipot, C.; Henin, J. *J. Chem. Phys.* **2005**, *123*, 244906/1.
(17) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
(18) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(19) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695.
(20) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737.
(21) Paci, E.; Ciccotti, G.; Ferrario, M.; Kapral, R. *Chem. Phys. Lett.* **1991**, *176*, 581.
(22) Ciccotti, G.; Ferrario, M. *Mol. Simul.* **2004**, *30*, 787.
(23) Grubmuller, H. *Phys. Rev. E* **1995**, *52*, 2893.
(24) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330.
(25) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703.
(26) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018.
(27) Martinez-Nunez, E.; Shalashilin, D. V. *J. Chem. Theory Comput.* **2006**, *2*, 912.
(28) Shalashilin, D. V.; Thompson, D. L. *J. Phys. Chem. A* **1997**, *101*, 961.
(29) In this article, we will often refer to "phase space" constraints. Strictly speaking, the examples that we provide in this work deal with geometrical constraints, so that "configuration space" is the more appropriate term. However, both AXD and BXD may be carried out in energy space,

which involves both coordinates and momenta, so that they are generalizable to phase space. To emphasize this generality, we often refer to phase space.

(30) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237.

(31) Miller James, A.; Klippenstein Stephen, J. *J Phys Chem A* **2006**, *110*, 10528.

(32) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. *J. Chem. Phys.* **2007**, *126*, 155101/1.

(33) Feige, M. J.; Paci, E. *J. Mol. Biol.* **2008**, *382*, 556.

(34) Paci, E.; Lindorff-Larsen, K.; Dobson, C. M.; Karplus, M.; Vendruscolo, M. *J. Mol. Biol.* **2005**, *352*, 495.

(35) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133.

(36) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins* **2002**, *46*, 24.

(37) Lourderaj, U.; Hase, W. L. *J. Phys. Chem. A* **2009**, *113*, 2236.

(38) Berendsen, H. J. C.; Marrink, S. J. *Pure Appl. Chem.* **1993**, *65*, 2513.

(39) Guo, Y.; Shalashilin, D. V.; Krouse, J. A.; Thompson, D. L. *J. Chem. Phys.* **1999**, *110*, 5521.

(40) Guo, Y.; Shalashilin, D. V.; Krouse, J. A.; Thompson, D. L. *J. Chem. Phys.* **1999**, *110*, 5514.

(41) Shalashilin, D. V.; Thompson, D. L. *J. Chem. Phys.* **1997**, *107*, 6204.

(42) Shalashilin, D. V.; Thompson, D. L. *ACS Symp. Ser.* **1997**, *678*, 81.

(43) Moroni, D.; van Erp, T. S.; Bolhuis, P. G. *Phys. Rev. E* **2005**, *71*, 5.

(44) Allen, R. J.; Frenkel, D.; ten Wolde, P. R. *J. Chem. Phys.* **2006**, *124*, 16.

(45) Borrero, E. E.; Escobedo, F. A. *J. Phys. Chem. B* **2009**, *113*, 6434.

(46) Escobedo, F. A.; Borrero, E. E.; Araque, J. C. *J. Phys.: Condens. Matter* **2009**, *21*, 23.

(47) Forney, M. W.; Janosi, L.; Kosztin, I. *Phys. Rev. E* **2008**, *78*, 9.

(48) Arnold, V. I. *Mathematical Methods of Classical Mechanics*; Springer-Verlag: New York, 1989.

(49) Lichtenberg, A. J.; Liberman, M. A. *Regular and Chaotic Dynamics*; Springer-Verlag: New York, 1992.

(50) Glowacki, D. R.; Marsden, S. P.; Pilling, M. J. *J. Am. Chem. Soc.* **2009**, *131*, 13896.

(51) Glowacki, D. R.; Reed, S. K.; Pilling, M. J.; Shalashilin, D. V.; Martinez-Nunez, E. *Phys. Chem. Chem. Phys.* **2009**, *11*, 963.

(52) Carpenter, B. K. *Ann. Rev. Phys. Chem.* **2005**, *56*, 57.

(53) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, *106*, 1737.

(54) Schwartz, S. D.; Schramm, V. L. *Nat. Chem. Biol.* **2009**, *5*, 552.

(55) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.

(56) Paci, E.; Cavalli, A.; Vendruscolo, M.; Caflisch, A. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 8217.

(57) Shirts, M. R.; Pande, V. S. *Phys. Rev. Lett.* **2001**, *86*, 4983.

(58) Garrett, B. C.; Truhlar, D. G. *Theory Appl. Comput. Chem.: First Forty Years* **2005**, 67.

(59) Truhlar, D. G.; Garrett, B. C. *Annu. Rev. Phys. Chem.* **1984**, *35*, 159.

(60) Georgievskii, Y.; Klippenstein, S. J. *J Phys Chem A* **2003**, *107*, 9776.

(61) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(62) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

(63) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.

(64) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

JP9074898