JCTC Journal of Chemical Theory and Computation

Using Selectively Applied Accelerated Molecular Dynamics to Enhance Free Energy Calculations

Jeff Wereszczynski* and J. Andrew McCammon

The Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

Abstract: Accelerated molecular dynamics (aMD) has been shown to enhance conformational space sampling relative to classical molecular dynamics; however, the exponential reweighting of aMD trajectories, which is necessary for the calculation of free energies relating to the classical system, is oftentimes problematic, especially for systems larger than small poly peptides. Here, we propose a method of accelerating only the degrees of freedom most pertinent to sampling, thereby reducing the total acceleration added to the system and improving the convergence of calculated ensemble averages, which we term *selective aMD*. Its application is highlighted in two biomolecular cases. First, the model system alanine dipeptide is simulated with classical MD, all-dihedral aMD, and selective aMD, and these results are compared to the infinite sampling limit as calculated with metadynamics. We show that both forms of aMD enhance the convergence of the underlying free energy landscape by 5-fold relative to classical MD; however, selective aMD can produce improved statistics over all-dihedral aMD due to the improved reweighting. Then we focus on the pharmaceutically relevant case of computing the free energy of the decoupling of oseltamivir in the active site of neuraminidase. Results show that selective aMD greatly reduces the cost of this alchemical free energy transformation, whereas all-dihedral aMD produces unreliable free energy estimates.

1. Introduction

Molecular dynamics (MD) simulations have become a crucial theoretical tool in advancing our understanding of the function of biological macromolecules.[1] Advances in algorithms[2] and computing power[3−5] continue to allow for simulations of increasingly larger systems on longer and longer time scales, permitting the direct observation of all-atom protein folding,[6,7] the observation of ion permeation through a transmembrane channel,[8] and the simulation of a complete virus.[9] Despite the remarkable progress that has been made in the field, simulation times still often fall far short of the miscosecond to millisecond time scales inherent in many biological processes. There have been several methodological advances which have aimed at simulating longer time scales within current computational power such as implicit solvation models,[10] multiple time stepping algorithms,[11] and improved treatment of long-range electrostat-

ics.[12] Sampling of phase space may also be enhanced through the deformation of the underlying potential energy surface,[13] as has been done in hyperdynamics,[14] puddle jumping,[15] conformational flooding,[16] and the local boost method[17] (to name only a few). Our group recently developed a method to enhance the crossing of barriers and the sampling of phase space termed accelerated molecular dynamics (aMD),[18] which has been shown to enhance sampling of biomolecular systems as in the conformational switching of Ras,[19] improve the agreement between experimental and calculated chemical shifts for IκBα,[20] and accelerate the calculation of pK_a values in lysozyme.[21]

An area of particular interest concerns using aMD simulations to calculate ensemble averages for physically relevant nonaccelerated systems. For systems in which the energy added to accelerate the system is low, trajectories may easily be reweighted; however, as the system size increases the boost energy required for significant acceleration increases, causing the exponential reweighting factor to produce

* Corresponding author e-mail: jmweresz@mccammon.ucsd.edu.

ensemble averages that are dominated by only a few configurations with high weight, thereby decreasing the precision of thermodynamic quantities such as free energy changes.[22] Here, we propose limiting the acceleration to the degrees of freedom most responsible for conformational changes, thereby reducing the energy added to a system to enhance sampling and resulting in improved reweighting statistics. This work is an extension of a previous study in which aMD was limited to the dihedrals in the backbone of a peptide substrate bound to cyclophilin;[23] however, here we demonstrate that accelerated dihedrals may contain atoms which are also contained in nonaccelerated torsions (that is, individual molecules may contain both accelerated and nonaccelerated torsions). Two examples are highlighted. First, we show that even in the case of the model system alanine dipeptide, selectively targeting dihedrals in the molecule's backbone results in similar acceleration levels while reducing the amount of energy added to the system. Then we turn our attention to the larger problem of calculating the binding energy of the clinically approved drug oseltamivir (marketed as Tamiflu by Roche Pharmaceuticals, Basel, Switzerland[24]) to the N1 flu protein neuraminidase.[25,26] Using free energy perturbation with a modification to the Bennett acceptance ratio (to account for reweighting),[27] we show that the computational cost required to accurately calculate the binding energy may be reduced by as much as 70% while maintaining a similar level of precision.

2. Theory

To enhance phase space sampling, the original aMD applies an additional potential only when the potential energy, $V(\mathbf{r})$, is below a specified criterion E to produce dynamics on the artificial landscape $V^*(r)$ such that

$$V^*(r) = \begin{cases} V(\mathbf{r}) & \text{if } V(\mathbf{r}) \geq E \\ V(\mathbf{r}) + \Delta V(\mathbf{r}) & \text{if } V(\mathbf{r}) < E \end{cases} \quad (1)$$

The form of the "boost" potential $\Delta V(\mathbf{r})$ is defined as

$$\Delta V(\mathbf{r}) = \frac{(E - V(\mathbf{r}))^2}{\alpha + (E - V(\mathbf{r}))} \quad (2)$$

The formalism of this boost potential has several practical advantages: it produces a potential energy surface with a smooth first derivative, it does not require the definition of a "reaction coordinate" along which to enhance sampling, it reflects the shape of the original potential, and it is relatively simple with only two adjustable parameters (E and α).

Simulation results generated with aMD may be reweighted by the exponential of the boost potential, $\exp(\beta\Delta V(\mathbf{r}))$, to recover theoretically exact thermodynamic properties for the physically relevant unaccelerated system. In practice, however, the exponential dependence of the reweighting hinders convergence as trajectory averages become dominated by a smaller subset of their snapshots as the range of boost potentials increases. While this does not severely affect small systems such as alanine dipeptide, it does prevent the use of aMD in accurate free energy calculation in larger biomolecular systems. To improve the reweighting statistics and enhance sampling, several variants of aMD have been

developed including "barrier lowering" aMD,[28] replica-exchange aMD,[29] and adaptive aMD (personal communication P. Markwick). In this paper we discuss an extension to aMD which may be incorporated into other aMD implementations, which is to selectively accelerate a user-defined subset of dihedrals most pertinent to sampling the relevant degrees of freedom, which we refer to as *selective aMD*. Selective aMD has the advantage that by only accelerating the degrees of freedom most important to sampling, lower overall boosts may be utilized to achieve a similar acceleration level, thus resulting in improved reweighting statistics. The idea of enhancing sampling along a user-defined manifold has been previously shown to improve the calculation of time-correlation functions for kinetics of multidimensional systems.[30]

2.1. Weighted Bennett Acceptance Ratio. Free energy perturbation (FEP) is a well-established technique used in free-energy calculations, specifically in the case of ligand binding and computational alchemy.[31,13] In FEP, a nonphysical energy pathway is constructed between two physical end states, for example, a ligand bound in the active site of an enzyme (which we denote as $\lambda = 0$) and an active site without the ligand ($\lambda = 1$). The path between these two states is divided into a series of "windows" in which the Hamiltonian is transformed from state 0 to 1. Traditionally, free energy differences between successive windows are estimated by exponentially averaging the instantaneous work of going between the states, and the overall free energy is a sum of free energy differences between windows.[32] Shirts et al. showed that the Bennett acceptance ratio (BAR) was superior to exponential averaging in producing asymptotically unbiased free energy estimates between two states that could improve precision by an order of magnitude.[27,33]

For a series of work functions between two states in which individual works do not each have the same weight (as in aMD), the derivation of a weighted BAR follows that in Shirts et al. with the exception that their eqs 5 and 6, the probability of a single measurement of the work W_i for the forward and reverse work functions, are modified to

$$P(F|W) = \frac{\exp(\beta\Delta V(\mathbf{r}))}{1 + \exp[-\beta(M + W_i - \Delta F)]} \quad (3)$$

$$P(R|W) = \frac{\exp(\beta\Delta V(\mathbf{r}))}{1 + \exp[\beta(M + W_i - \Delta F)]} \quad (4)$$

where the constant M is redefined as

$$M = kT \ln\left(\frac{\sum_{i=0}^{n_F} \exp(\beta\Delta V(\mathbf{r})_i)}{\sum_{j=0}^{n_R} \exp(\beta\Delta V(\mathbf{r})_j)}\right) \quad (5)$$

Therefore, the value of ΔF that solves

$$\sum_{i=0}^{n_F} \frac{\exp(\beta\Delta V(\mathbf{r})_i)}{1 + \exp[-\beta(M + W_i - \Delta F)]} =$$
$$\sum_{j=0}^{n_R} \frac{\exp(\beta\Delta V(\mathbf{r})_j)}{1 + \exp[-\beta(M + W_j - \Delta F)]} \quad (6)$$

is the optimal free energy estimate between adjacent windows. It has recently been shown that reweighting of states in BAR to account for non-Boltzmann sampling may have practical advantages outside of aMD simulations.[34]

3. Computational Details

Molecular Dynamics Details. MD simulations were performed with the MD package Desmond (version 2.2) developed by D. E. Shaw Research.[35] Both systems were built, solvated, and ionized with Schrödinger's Maestro modeling suite such that there was a minimum of 12 Å of TIP3P[36] water buffer between the macromolecule and the periodic boundary and an ionic concentration of ~150 mM NaCl was present. The CHARMM22 force field with the CMAP correction was utilized (except where noted below in the neuraminidase calculations).[37] Following 10 000 steps of minimization, systems were continuously heated to 300 K over 1.5 ps. All simulations used the Martyna−Tobias−Klein constant pressure and temperature algorithms (a combination of Nosè−Hoover constant temperature and piston constant pressure algorithms)[38,39] with a reference temperature and pressure of 300 K and 1.01425 bar, respectively. Short-range nonbonded interactions were truncated at 12 Å, while long-range electrostatics were calculated with a particle-mesh Ewald algorithm using a sixth-order B-spline for interpolation and a grid spacing of <1 Å in each dimension.[40] A time step of 2 fs was employed, and the M-SHAKE algorithm was used for constraining all hydrogen-containing bonds.[41] A plugin was written for Desmond to perform aMD calculations on specified dihedrals.

Alanine Dipeptide. An alanine dipeptide molecule based on a model compound was solvated in a $(27 \text{ Å})^3$ box using Maestro and equilibrated for 5 ns following heating. Two sets of 50 ns aMD trajectories were run, one in which all dihedrals were accelerated with aMD and one in which only the two dihedrals defined as ϕ and ψ were accelerated (selective aMD, see Figure 1). For each setup, 16 simulations spanning the parameter space of E and α were run to optimize these parameters. For the all-dihedral simulations, parameters of $E = 8$ and $\alpha = 4$ had an optimal fit to our metric χ (as discussed below), whereas for the selective aMD $E = 1$ and $\alpha = 0.75$ were optimal. Additionally, a 250 ns classical MD simulation was performed.

To determine the accuracy of the aMD results, well-tempered metadynamics was performed to calculate the underlying two-dimensional free energy landscape in ϕ/ψ space.[42] In well-tempered metadynamics, the height of a Gaussian centered at position x is proportional to the Boltzmann weight of the metadynamics potential already present at x, that is the added Gaussian has a maximum value of $\omega_0 \cdot \exp(-V_t(x)/k\Delta T)$, with ω_0 being the initial Gaussian height, $-V_t(x)$ the metadynamics potential at x, and $k\Delta T$ a
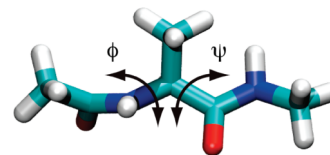

Figure 1. Conformation of the model system alanine dipeptide can be expressed by the angles of the two torsions ϕ and ψ.

user-defined energy which limits the explored energy range. We performed a 50 ns simulation in which gaussians were added every 0.2 ps with a width of 0.1 radians and a height determined by $\omega_0 = 0.02$ kcal/mol and $k\Delta T = 2.4$ kcal/mol. To quantitate convergence we define a metric χ as follows

$$\chi = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{U=.6,.7,\ldots}^{3.0} \begin{cases} P_{i,U}^{\text{aMD}}/P_{i,U}^{\text{exact}} & \text{if } P_{i,U}^{\text{exact}} > P_{i,U}^{\text{aMD}} \\ P_{i,U}^{\text{exact}}/P_{i,U}^{\text{aMD}} & \text{if } P_{i,U}^{\text{exact}} < P_{i,U}^{\text{aMD}} \end{cases} \quad (7)$$

For a given well (defined as i in Figure 2f) this metric calculates the ratio of the population of states below the energy U ($P_{i,U}^{\text{aMD}}$) with that expected from the metadynamics results ($P_{i,U}^{\text{exact}}$) and averages this over M energy values of U from 0.6 to 3 kcal/mol (in increments of 0.1 kcal/mol) for each of the $N = 3$ wells. If a well has a population greater than that expected from the metadynamics result, the inverse of the ratio is taken to equally account for over and under sampling of the well. This metric has several advantageous features: by averaging over multiple energy levels it selects for smooth population densities (as are observed in cMD and metadynamics but may not result when the trajectory is reweighted with aMD), it treats over and under sampling a well as equally poor, and it equally weights all three of the main energy wells. $\chi = 1$ is considered an ideal reproduction of the exact population.

N1 Neuraminidase. A monomer of neuraminidase bound to the inhibitor oseltamivir taken from the 2HU0 crystal structure was solvated in an approximately $(70 \text{ Å})^3$ box. Following the minimization and heating protocol outlined above, the system was equilibrated for 5 ns with the AMBER99SB force field[43] before protein and water parameters were changed to the CHARMM22 force field and further equilibrated for 1 ns. Parameters for oseltamivir were previously developed for use with the AMBER99SB force field[44] and maintained throughout the CHARMM simulations. The switch to the CHARMM force field was performed after testing of aMD free energy calculations revealed that increased acceleration levels in the AMBER force field tended to disturb the electrostatic components of the free energy calculation; therefore, this hybrid AMBER (for the ligand) and CHARMM (for the remainder of the system) force field was utilized. While the authors concede this may produce incorrect absolute binding energies, the goal of this paper is to study convergence of accelerated free energy calculations to results obtained from unaccelerated (and longer) calculations.

Alchemical free energy calculations for the decoupling of the ligand in the protein's active site were performed using 21 windows in which the electrostatics were decoupled over
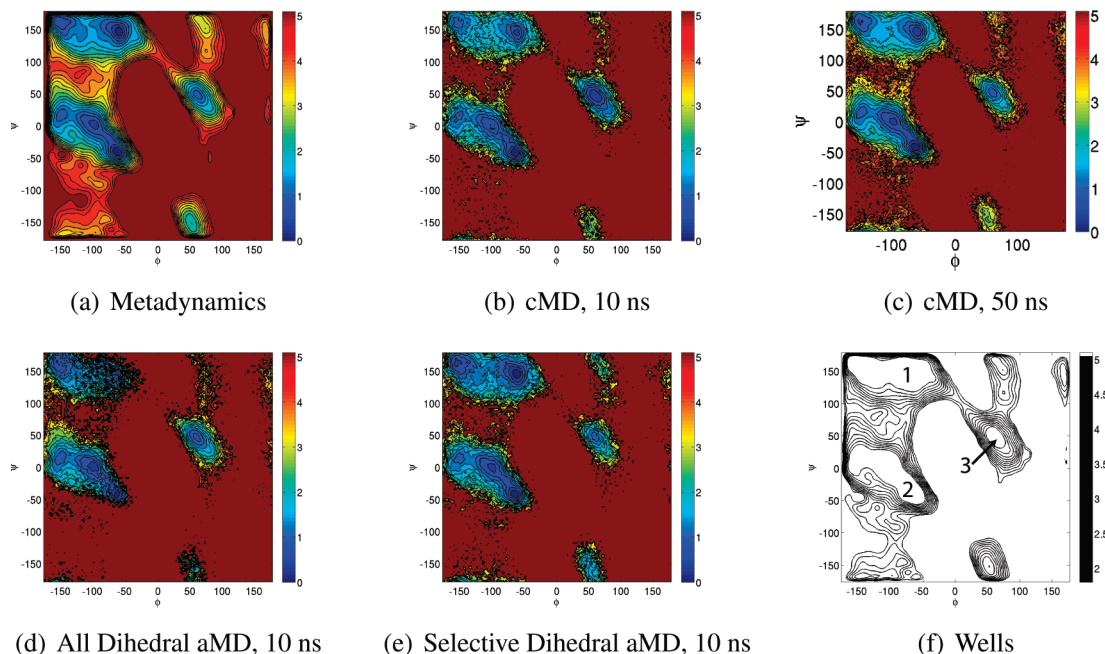
(a) Metadynamics

(b) cMD, 10 ns

(c) cMD, 50 ns

(d) All Dihedral aMD, 10 ns

(e) Selective Dihedral aMD, 10 ns

(f) Wells

Figure 2. Free energy landscapes for alanine dipeptide. (a) Metadynamics results representing the expected results for infinite sampling of the system. Classical MD simulations visited all three energy wells on the 10 ns time scale (b); however, the statistics are poor (as shown in Table 1). Sampling is improved on the 50 ns time scale (c). (d) All dihedral aMD, showing that a sampling of all three energy wells is improved relative to the cMD results. (e) Selective dihedral aMD, showing the best agreement to the metadynamics results on the 10 ns time scale. (f) Wells labeled for discussion purposes.

Table 1. Free Energy Statistics for Alanine Dipeptide from cMD, All-Dihedral aMD, and Selective aMD Simulations Compared to the Metadynamics Results Which Effectively Represent the Infinite Sampling Limit[a]

		well 1		well 2		well 3	
time	simulation type	energy	population	energy	population	energy	population
	metadynamics	0.065	0.406	0.000	0.429	1.010	0.049
10 ns	classical	0.485 (+.421)	0.249 (61%)	0.000 (+.000)	0.427 (99%)	0.301 (−0.706)	0.191 (392%)
	all dihedral aMD	0.000 (−0.065)	0.337 (83%)	0.085 (+.085)	0.415 (97%)	0.767 (−0.240)	0.078 (160%)
	selective aMD	0.008 (−0.057)	0.422 (104%)	0.000 (+.000)	0.369 (86%)	1.010 (−0.004)	0.041 (84%)
50 ns	classical	0.350 (+.285)	0.328 (81%)	0.000 (+.000)	0.475 (111%)	1.004 (−0.002)	0.058 (120%)
	all dihedral aMD	0.053 (−0.012)	0.372 (92%)	0.000 (+.000)	0.426 (99%)	1.003 (−0.004)	0.052 (106%)
	selective aMD	0.137 (+.072)	0.381 (94%)	0.000 (+.000)	0.422 (98%)	1.038 (+.033)	0.048 (98%)
250 ns	classical	0.232 (+.168)	0.357 (88%)	0.000 (+.000)	0.446 (104%)	0.994 (−0.012)	0.055 (112%)

[a] For each of the three wells (Figure 2f) the minimum energy relative to the global minimum is calculated, as is the population of states within 1.8 kcal/mol of that minimum. Results show that both aMD forms converge on the order of five times as fast as the cMD simulations.

10 windows followed by the Lennard−Jones interactions decoupled over 10 windows with a softcore potential using $\alpha = 0.5$[45] (as in a previous study[46]). Free energies were calculated using BAR, with the modified BAR formulas described above used in the aMD calculations. A positional restraint of .8 kcal/mol was placed on a central carbon atom in oseltamivir to prevent the ligand from sampling nonactive site portions of the simulation box,[47] and calculations were performed for cMD, all-dihedral aMD (with $E = 2600$ and $\alpha = 400$), and selective aMD (with $E = 13$ and $\alpha = 2$). For cMD calculations, three sets of windows were run (as has been shown to improve calculated free energies[48]), each with the same initial coordinates but different velocities for 5 ns per window, whereas for the selective aMD the same three sets of windows were run with 200 ps of all-dihedral aMD (to quickly equilibrate the whole protein) followed by 1.5 ns of selective aMD per window. Note that the times indicated in the text *include* the simulation time spent in all-dihedral aMD; thus, a time of 500 ps represents 200 ps of

all-dihedral and 300 ps of selective aMD. One set of FEP calculations was performed for the all-dihedral case for 1.75 ns window to illustrate the futility of using standard aMD in large-scale biomolecular FEP calculations.

The choice of dihedrals to accelerate was based on previous work in which the tetramer was simulated for 100 ns.[46] Acceleration was applied to those dihedrals which contained only heavy atoms, were in residues that had a heavy atom within 5 Å of the oseltamivir in the crystal structure, and had a multivariate distribution for a total of 29 dihedrals.

Work functions were decorrelated based on the statistical inefficiency using code provided by Shirts and Chodera.[49] For each BAR calculation, a bootstrap analysis was performed (with 50 independent calculations) for an error σ_B, which was combined with the variance of the three means (σ_V) to calculate an overall error estimate for the free energy by

$$\sigma = \sqrt{\sigma_V + \sum_{i=1}^{3} \sigma_B / 3} \qquad (8)$$

Results

Alanine Dipeptide. The free energy landscape of alanine dipeptide can be described by the rotation of two torsional angles, ϕ and ψ, making it an ideal model system for methodological development that has been extensively studied (Figure 1). The metadynamics, or "infinite sampling limit", results (Figure 2a) show three distinct energy wells which we label for further discussion in Figure 2f. The energy barrier between wells 1 and 2 is relatively low, and classical MD (cMD) simulations sample both sets of configurations on the 10 ns time scale (Figure 2b). Well 3 is substantially oversampled, which we attribute to the system becoming trapped in this state due to the higher energy barrier, thereby discouraging transitions to and from well 3. With 50 ns of simulation the sampling of all three wells are improved (Figure 2c). Further detail is shown in Table 1, which compares the minimum energy of each well and probability of all states within 1.8 kcal/mol of that minimum to the theoretically exact answer derived from metadynamics. The 10 ns cMD shows good agreement for well 2 (it is identified as the global minimum and the population at 3kT is nearly identical to the metadynamics results); however, well 1 is undersampled by 39% whereas well 3 is oversampled by 292%, and the minimum energies are incorrect by .4 and .7 kcal/mol, respectively. With 50 ns of simulation time the sampling improves such that errors in the populations range from 11% to 20%, and by 250 ns of sampling the statistics agree much better with the metadynamics results for all three wells, although the populations of wells 1 and 3 are still off by >10%.

For comparison, the all-dihedral and selective aMD simulations sampled all three energy wells on the 10 ns time scale (Figure 2d and 2e). Free energy statistics indicate a maximum error in the minimum well energy estimate of 0.24 kcal/mol in the all-dihedral case and 0.06 kcal/mol in the selective aMD simulation, whereas the greatest disagreement in well populations was an undersampling of well 1 by 17% in the all-dihedral aMD and an undersampling of well 3 by 16% in the selective aMD. Extension of the simulations to 50 ns results in further improved statistics of the well populations, with well 1 only being undersampled by 8% in the all-dihedral simulation and by 6% in the selective aMD.

To further examine the convergence of the free energy statistics we defined the parameter χ to quantitate the difference in the two-dimensional energy profiles (as discussed in the methods) which has the property of a value of 1 representing ideal sampling of the wells as compared to the metadynamics results. In Figure 3 we compare the time course of this parameter between the cMD and the aMD simulations (note the different time scales for the two sets of simulations). The scores for both aMD are similar to those for cMD simulations of five times the length, with cMD simulations requiring 200 ns before consistently having values above 0.9, whereas the aMD simulations pass this value at 44 and 35 ns for all-dihedral and selective aMD, respectively.
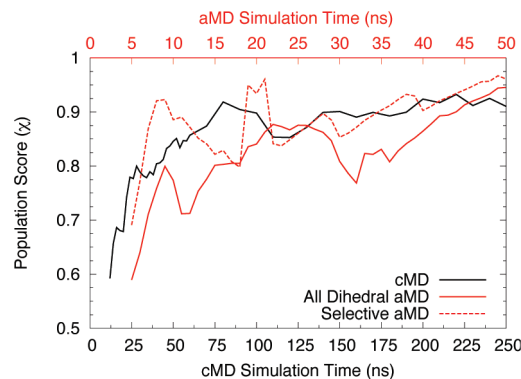

Figure 3. Comparison of the time evolution of our order parameter χ. Classical MD results (shown in black) correspond to the lower *x* axis, which is five times longer than that of the upper *x* axis, which corresponds to both aMD simulations. Despite this 5-fold difference in time scales, all three simulations show similar convergence patterns, with the lower weights in the selective dihedral slightly improving the statistics relative to the all-dihedral aMD.

A comparison of the boosts applied throughout the aMD simulations (Figure 4a) shows that higher boost potentials are applied throughout the all-dihedral simulation than in the selective aMD. For the all-dihedral case, ΔV varies from 0 to 4.2 kcal/mol, with an average value of 0.45 kcal/mol, whereas the selective aMD has a range of 0 to 0.54 kcal/mol with an average of 0.11 kcal/mol. This decrease in the applied boosts has a significant impact on reweighting as the maximum weight relative to an unaccelerated state is reduced from 1097 to 2.46. In Figure 4b the amount of total weight recovered from a simulation is plotted against the percentage of frames that contribute to that weight. In the case of all-dihedral aMD, very few frames contribute a substantial portion of the reweighting, 50% of the total weight comes from 4.8% of the trajectory, whereas 90% of the weight comes from 53.7% of the trajectory. This results in almost one-half the sampling (46.3%) contributing very little (less than 10%) to the calculated ensemble averages. In the selective aMD case the lower boosts result in more uniform weights, 50% of the weight comes from 37.9% of the trajectory and 90% from 87.6% of the trajectory (for comparison, in cMD configurations in the trajectory are uniformly weighted, so 50% of the weight comes from 50% of the trajectory). This increased reweighting efficiency improves the recovered statistics as high-boost configurations tend to dominate the ensemble average. For example, wells 1 and 2 appear significantly smoother in the energy landscape of the selective aMD relative to the all-dihedral aMD (Figure 2e and 2d) and χ is consistently higher in the selective case, both of which can be attributed to smoother statistics in the reweighted energy profiles.

N1 Neuraminidase. The binding of oseltamivir to neuraminidase is an example of how highly accurate free energy calculations may be employed in the study and development of novel pharmaceutical compounds. In order to validate our aMD simulations of the decoupling of oseltamivir in the N1 active site, we performed extensive sampling at each step of the alchemical transformation with 5 ns of cMD simulation for each of the 21 λ values, which
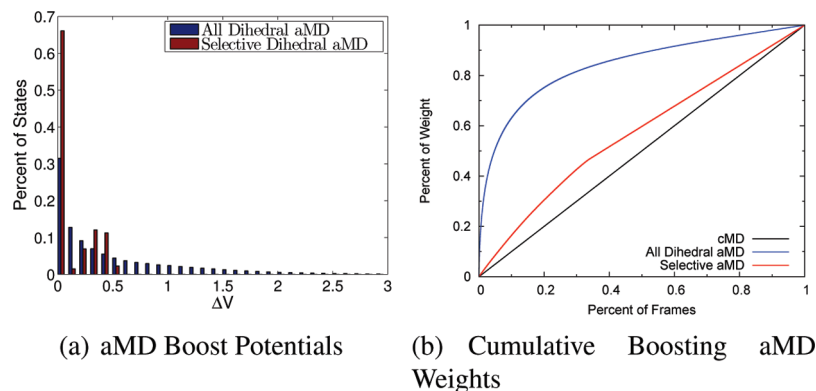
(a) aMD Boost Potentials   (b) Cumulative Boosting aMD Weights

Figure 4. (a) Distribution of boost potentials applied during the 50 ns simulations, showing an increased range of ΔV values throughout the all-dihedral aMD simulation relative to the selective dihedral aMD. (b) Percent of the total weight from the simulations relative to the percent of frames that contributed this weight. The lower ranges of boosts in the selective aMD produce more uniform sampling relative to the all-dihedral aMD, which tends to be dominated by a small subset of the configurations.
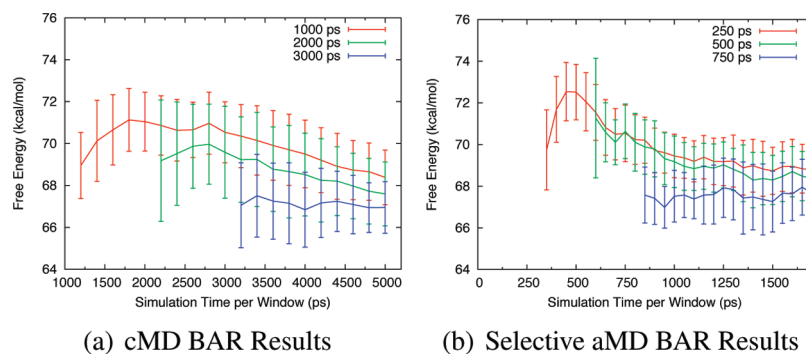

(a) cMD BAR Results   (b) Selective aMD BAR Results

Figure 5. Free energies for the decoupling of oseltamivir from the neuraminidase active site as computed by alchemical transformation with classical MD (a) and selectively applied aMD (b). For each set of simulations three possible equilibration times were chosen; however, for only the longest equilibration period does the calculated free energy remain stable with increased sampling. The aMD results converged to values comparable to the cMD results with significantly less simulation time. The time scale on the *x* axis includes the equilibration time for which free energy values are not calculated.

was repeated three times (with different initial velocities). An equilibration period is typically discarded from the BAR calculations, the length of which is determined by several factors, including molecular rigidity, the slow motions of loops which are relevant to free energy differences, and the amount of computational time available.[50] In Figure 5a we show the time evolution of the mean free energies for three equilibration times, 1000, 2000, and 3000 ps, along with the associated errors. The 3000 ps of equilibration curve is the only one which does not have a mean that changes appreciably with increased sampling, suggesting that for this initial configuration of this complex a full 3 ns per window is required for each of the λ windows to equilibrate to their new Hamiltonians. The free energy of 66.9 ± 1.2 kcal/mol is calculated by using 3 ns for equilibration and the remaining 2 ns for sampling, as shown in Table 2. Note that this is not the free energy of binding in solution; rather it is only one leg of the thermodynamic cycle required for that calculation. Examination of individual BAR runs shows that increased sampling decreases both the variance between the three runs and the bootstrap errors associated with each (Figure S1, Supporting Information).

BAR results from simulations using selective aMD (with the first 200 ps utilizing all-dihedral aMD) are presented in Figure 5b. As in the cMD calculations, we have chosen three

Table 2. Free Energies for the Decoupling of Oseltamivir from the Neuraminidase Active Site As Calculated with cMD and Selective aMD for Varying Equilibration and Sampling Periods

simulation type	equilibration time	sampling time	free energy (kcal/mol)
cMD	1 ns	1 ns	71.0 ± 1.4
		2 ns	70.5 ± 1.5
		3 ns	69.5 ± 1.6
		4 ns	68.4 ± 1.3
	2 ns	1 ns	69.6 ± 2.2
		2 ns	68.5 ± 2.0
		3 ns	67.6 ± 1.5
	3 ns	1 ns	66.8 ± 1.8
		2 ns	66.9 ± 1.2
aMD	250 ps	500 ps	70.5 ± 1.4
		750 ps	69.4 ± 1.0
		1000 ps	69.2 ± 1.2
		1250 ps	68.7 ± 1.1
	500 ps	500 ps	69.2 ± 1.2
		750 ps	69.0 ± 1.1
		1000 ps	68.3 ± 1.2
	750 ps	500 ps	67.9 ± 1.4
		750 ps	67.3 ± 1.5

equilibration times, 250, 500, and 750 ps, and the mean value of the free energy does not remain constant with increased sampling unless the longest of these equilibration times is discarded. With 750 ps of equilibration and 750 ps of sampling

sing Selectively Applied Accelerated Molecular Dynamics

J. Chem. Theory Comput., Vol. 6, No. 11, 2010 **3291**

we obtain a free energy value of 67.3 ± 1.5 kcal/mol, identical (within error) to that from the longer cMD calculations. Results from individual aMD runs are shown in Figure S2, Supporting Information. A comparison of the time evolutions of the BAR results shows similar behavior for the aMD and cMD free energies (with the aMD on shorter time scales) for the short, medium, and long equilibration times (Table 2 and Figure S3, Supporting Information). For short equilibration periods (cMD, 1000 ps; aMD, 250 ps) the free energy is initially overestimated, and while it approaches the values obtained for longer equilibration, the bias introduced in this nonequilibrated period results in free energies ~1.5 kcal/mol too high. Medium length equilibration periods (cMD, 200 ps; aMD, 500 ps) suffer from this effect as well but are not quite as biased, whereas for long equilibration times (cMD, 3000 ps; aMD, 750 ps) the calculated free energies remain stable (within error) with increased sampling.

As a comparison, we also performed BAR calculations on a single set of windows run for 1750 ps with all-dihedral aMD (Figure S4, Supporting Information). The much larger range of weights resulted in very few configurations contributing to the BAR results and poor free energy estimates. For example, with 250 ps of sampling and 1250 ps of equilibration a free energy of -15.0 ± 17.0 kcal/mol was computed, whereas 750 ps of equilibration and 750 ps of sampling resulted in a free energy of decoupling of 36.9 ± 17.3 kcal/mol. Both of these values are wildly inaccurate compared to the values of 66.9 ± 1.2 and 67.3 ± 1.5 kcal·mol obtained from cMD and selectively applied aMD.

5. Concluding Discussions

The method of accelerated molecular dynamics has been well established as a means of enhancing phase space sampling with minimal computational cost; however, the exponential reweighting required for the recovery of ensemble averages in the unaccelerated case introduces excessive noise such that it is often difficult, if not impossible, to recover accurate ensemble averages. Even in the case of the well-studied system alanine dipeptide this becomes evident. For example, in Figure 2d the free energy profile of well 1 appears discontinuous in the region of $(\phi, \psi) = (-50, 150)$, due in large part to the fact that several of the conformations visited have weights $1-3$ orders of magnitude below that of the maximum weighted conformation. In contrast, by limiting the acceleration to only those dihedrals which are most pertinent to phase space sampling (in this case, ϕ and ψ) the maximum weight was reduced from 1097 to 2.46, which not only increased the smoothness of the recovered free energy profile (Figure 2e) but also moderately improved the agreement to the infinite sampling limit (as calculated with metadynamics) results in Table 1 and Figure 3. However, both aMD formalisms showed approximately a factor of 5-fold increase in efficiency relative to cMD for our order parameter χ.

Extension of this idea to the pharmaceutically relevant case of oseltamivir binding to neuraminidase shows the expensive free energy calculation of the decoupling of the ligand in the protein's active site may be reduced by up to 70% (from 5 to 1.5 ns/window) without a loss in precision. While this may not be crucial in the case of studying the binding of

only a single ligand to a protein, one could imagine that in the lead optimization stage of a drug-design effort, when highly accurate binding energies are necessary, the ability to examine three times the number of possible compounds at little extra computational cost may be highly desirable. In the case presented here, the accelerated dihedrals were chosen based upon extensive prior MD simulations; however, if this data were not available one could choose the accelerated dihedrals by residue type (accelerating dihedrals in residues with highly mobile side chains such as arginine and not accelerating dihedrals in aromatic rings), atom types (non-hydrogen containing), and proximity to the ligand. Additionally, in some cases where the ligand is bulky and has multiple torsions with high-energy barriers between local minima, one could accelerate dihedrals in the ligand molecules themselves, as was done in the case of cyclophilin.[23]

Selective aMD may easily be incorporated into other free energy algorithms. For example, the methods of one-step perturbation and envelope distribution sampling provide techniques for effectively calculating the binding of several ligands to a protein with a single extended MD simulation.[51,52] They have, however, not been extensively utilized due to the fact that, depending on the system being studied, they may require simulations on the microseconds time scale.[53] Therefore, a reduction of the computational cost of 3- to 5-fold (as observed in this study) could reduce the necessary simulation length from the highly expensive $1-2$ μs range into the manageable 200 ns time scale. These methods highlight that sampling of the partition function is, in general, a slow process requiring extensive calculations; therefore, methods such as selective aMD, which can enhance sampling of the relevant portions of phase space while not introducing excessive noise into the calculations, may prove useful in future applications.

Supporting Information Available: Plots of the free energies for oseltamivir decoupling as a function of sampling time for individual cMD and aMD simulations; comparisons between these cMD and aMD runs; and free energies calculated from a single all-dihedral aMD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(3) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *J. Comput. Chem.* **2007**, *28*, 2618–2640.

(4) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. *J. Comput. Chem.* **2009**, *30*, 864–872.

(5) Shaw, D. E.; et al. *Commun. ACM* **2008**, *51*, 91–97.

(6) García, A. E.; Onuchic, J. N. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(7) Ensign, D. L.; Kasson, P. M.; Pande, V. S. *J. Mol. Biol.* **2007**, *374*, 806–816.

(8) Jensen, M. O.; Borhani, D. W.; Lindorff-Larsen, K.; Maragakis, P.; Jogini, V.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *Proc. Nat. Acad. Sci. U.S.A.* **2010**, *107*, 5833–5838.

(9) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437–449.

(10) Chen, J. H.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148.

(11) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(12) Shan, Y. B.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2005**, *122*, 054101.

(13) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 1569–1582.

(14) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908–3911.

(15) Rahman, J. A.; Tully, J. C. *Chem. Phys.* **2002**, *285*, 277–287.

(16) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893–2906.

(17) Wang, J. C.; Pal, S.; Fichthorn, K. A. *Phys. Rev. B* **2001**, *63*, 085403.

(18) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(19) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. *PLoS Comput. Biol.* **2009**, *5*, e100325.

(20) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 1220–1221.

(21) Williams, S. L.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 560–568.

(22) Shen, T. Y.; Hamelberg, D. *J. Chem. Phys.* **2008**, *129*, 034103.

(23) Hamelberg, D.; McCammon, A. *J. Am. Chem. Soc.* **2009**, *131*, 147–152.

(24) Lew, W.; Chen, X. W.; Kim, C. U. *Curr. Med. Chem.* **2000**, *7*, 663–672.

(25) Landon, M. R.; Amaro, R. E.; Baron, R.; Ngan, C. H.; Ozonoff, D.; McCammon, J. A.; Vajda, S. *Chem. Biol. Drug Des.* **2008**, *71*, 106–116.

(26) Cheng, L. S.; Amaro, R. E.; Xu, D.; Li, W. W.; Arzberger, P. W.; McCammon, J. A. *J. Med. Chem.* **2008**, *51*, 3878–3894.

(27) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.

(28) de Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *J. Chem. Theory Comput.* **2008**, *4*, 1516–1525.

(29) Fajer, M.; Hamelberg, D.; McCammon, J. A. *J. Chem. Theory Comput.* **2008**, *4*, 1565–1569.

(30) MacFadyen, J.; Wereszczynski, J.; Andricioaei, I. *J. Chem. Phys.* **2008**, *128*, 114112.

(31) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(32) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

(33) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.

(34) König, G. ; Boresch, S. *J. Comput. Chem.* **2010**, in press.

(35) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC'06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*; ACM Press: New York, 2006; p 84.

(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(37) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(38) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(39) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(40) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(41) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501–508.

(42) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(43) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(44) Amaro, R. E.; Minh, D. D. L.; Cheng, L. S.; Lindstrom, W. M.; Olson, A. J.; Lin, J. H.; Li, W. W.; McCammon, J. A. *J. Am. Chem. Soc.* **2007**, *129*, 7764–77677655.

(45) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Vangunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(46) Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2523–2532.

(47) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(48) Lawrenz, M.; Baron, R.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, *5*, 1106–1116.

(49) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

(50) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 2618–2628.

(51) Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1996**, *100*, 9485–9494.

(52) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2008**, *128*, 174112.

(53) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Nat. Acad. Sci. U.S.A.* **2005**, *102*, 6750–6754.