JCTC Journal of Chemical Theory and Computation

Coupling Accelerated Molecular Dynamics Methods with Thermodynamic Integration Simulations

César Augusto F. de Oliveira,* Donald Hamelberg, and J. Andrew McCammon

Howard Hughes Medical Institute, Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received May 12, 2008

Abstract: In this work we propose a straightforward and efficient approach to improve accuracy and convergence of free energy simulations in condensed-phase systems. We also introduce a new accelerated Molecular Dynamics (MD) approach in which molecular conformational transitions are accelerated by lowering the energy barriers while the potential surfaces near the minima are left unchanged. All free energy calculations were performed on the propane-to-propane model system. The accuracy of free energy simulations was significantly improved when sampling of internal degrees of freedom of solute was enhanced. However, accurate and converged results were only achieved when the solvent interactions were taken into account in the accelerated MD approaches. The analysis of the distribution of boost potential along the free energy simulations showed that the new accelerated MD approach samples efficiently both low- and high-energy regions of the potential surface. Since this approach also maintains substantial populations in regions near the minima, the statistics are not compromised in the thermodynamic integration calculations, and, as a result, the ensemble average can be recovered.

Introduction

Free energy is probably the most important quantity in thermodynamics and one of the central topics in biophysics.[1,2] Nevertheless, for many relevant systems with local minimum energy configurations separated by energy barriers, efficient and accurate calculation of this property is still a big challenge in computational chemistry. Free energy differences between different states can be calculated through Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) methods.[3−10] Since the first application of the methodology to the calculation of the relative free energies of ligand binding and solvation of the organic molecules methanol and ethane,[1,8] FEP and TI have been widely used to study a wide range of processes such as solvation, phase transitions, ligand binding, and protein−protein interactions, just to name a few.[5,11−14] These methods, which are firmly rooted in statistical mechanics, are usually combined with molecular dynamics (MD) or Monte Carlo (MC) simulations.[15,16]

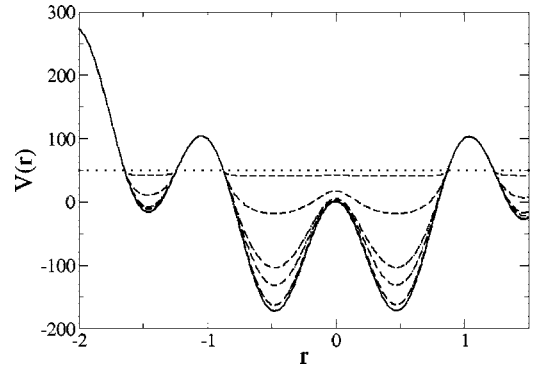

Figure 1. Schematic representation of a hypothetical true (solid line) and modified (dashed line) potential energy function with different values of α. The modified potential (generated with eq 1) converges to the true potential at large values of α. The dotted line corresponds to the boost energy *E*.

The data obtained from these simulations allows us to quantitatively evaluate free energy changes and understand, at molecular level, the structural and energetic factors governing the process.

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

10.1021/ct800160q CCC: $40.75
Published on Web 08/13/2008

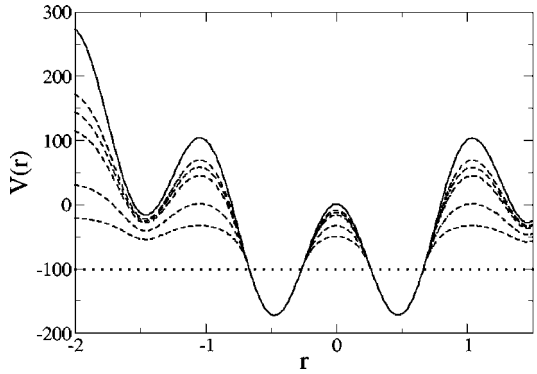

Figure 2. Schematic representation of a hypothetical true (solid line) and modified (dashed line) potential energy function with different values of α. The modified potential (generated with eq 3) converges to the true potential at large values of α. The dotted line corresponds to the boost energy *E*.

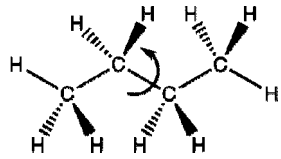

Figure 3. Butane molecule.

However, to obtain accurate free energy values, major issues like free energy convergence and conformational sampling still need to be addressed. Although these topics have been mainly discussed as independent issues, convergence and the amount of sampling are strictly connected.[17] For instance, for processes that involve large conformational changes and reorganization of solvent, poor sampling can trap the system in local minima and, as a consequence, lead to apparent but false convergence. In other words, in these cases the calculated free energy might correspond to pseudoconverged values obtained from trapped local conformations. As we will show in this paper, even for a very simple system, like propane-to-propane transformation, independent free energy calculations carried out with conventional MD simulation may not be able to reproduce accurately the correct free energy difference, though the simulations may show apparently converged values. Quantitative prediction of free energy change is only obtained when configuration sampling is efficiently improved.

A large number of techniques have been introduced to enhance sampling over configuration space.[18−32] A straightforward way of modifying the potential energy surface to enhance sampling has been proposed by Hamelberg et al.[33] This approach, which is based on earlier work of Voter,[34,35] has proved to be efficient in accelerating not only conformational transitions[36−39] but also millisecond time scale motions of a protein in explicit water.[40]

In this work we propose a simple and efficient approach to improve accuracy and convergence of free energy simulations in condensed-phase systems. The main idea is to integrate the accelerated MD approach with free energy simulations. Although the formulation and the results presented here were obtained by coupling TI with the accelerated MD method (aMD), the procedure can be easily extended to the FEP approach. To check convergence and accuracy of the TI simulations, all calculations were performed on the propane-to-propane system. This system was chosen because i) the correct free energy result is rigorously equal to zero and ii) similar "zero-free energy change" systems have been used before as model systems to compare the efficiency and convergence of different approaches to free energy calculations.[41,42]

Theory

In order to enhance sampling by increasing the escape rate from potential energy wells, the accelerated MD approach modifies the energy landscape by adding a boost potential, $\Delta V(r)$, to the original potential surface every time $V(r)$ is below a predefined energy level E (Figure 1). In other words,

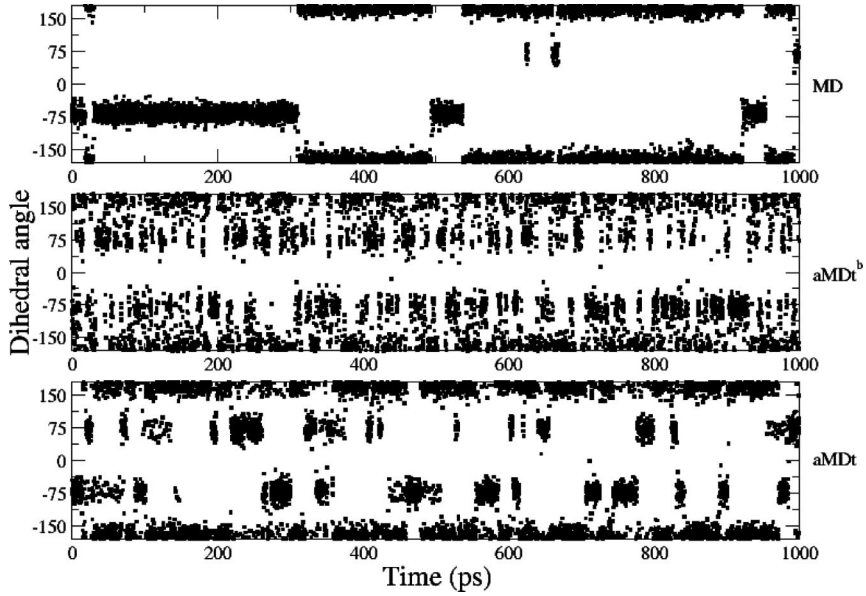

Figure 4. Plots of dihedral angle of the butane molecule, as defined in Figure 3, sampled with normal MD, aMDt^b, and aMDt approaches.

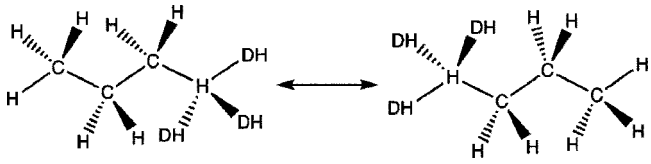

Figure 5. Propane-to-propane transformation. DH stands for dummy atoms.

$V^*(r) = V(r) + \Delta V(r)$. In the Hamelberg et al.[33] implementation, $\Delta V(r)$ is given by

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + (E - V(r))}, & V(r) < E \end{cases} \quad (1)$$

where α modulates the depth and the local roughness of the energy basins in the modified potential. Since the torsional potential governs the rate of sampling of biomolecular rotameric states, the boost potential has been largely applied to the torsional term of the potential energy function. This approach, which will be referred to as aMDt, has been successfully applied to study several biological systems and processes.[37−39,43,44]

More recently, Hamelberg et al. introduced a dual boost approach in order to efficiently sample both the torsional degrees of freedom and the diffusive motions.[36] In this implementation, two boost potentials are applied separately to the potential energy. While the first one is applied only to the torsional terms, the second one is added to the total potential energy (aMDtT). The modified potential is given by

$$V^*(r) = \{V_0(r) + [V_t(r) + \Delta V_t(r)]\} + \Delta V_T(r) \quad (2)$$

where $\Delta V_t(r)$ and $\Delta V_T(r)$ are the boost potentials applied to the torsional terms $V_t(r)$ and the total potential $V_T(r)$. $V_0(r)$ is the potential energy excluding contribution from torsional terms. Both boost potentials are defined according to eq 1. Here $V_T(r)$ is defined as $V_T(r) = V_0(r) + V_t(r) + \Delta V_t(r)$.

The correct canonical averages of an observable, calculated from configurations sampled on the modified potential energy surface, is then fully recovered from the accelerated MD simulations by reweighting each point in the configuration space by $\exp\{\beta[\Delta V(r)]\}$. In the dual boost approach, the boost factor is given by $\exp\{\beta[\Delta V_t(r)+\Delta V_T(r)]\}$.

New Accelerated MD Approach. In this work, a third approach is introduced in which molecular conformational transitions are accelerated by lowering the energy barriers, while the potential surfaces near the minima are left unchanged. The idea behind this approach has been used before by Darve et al. to calculate free energies by applying a scale-force molecular dynamics algorithm.[27]

Owing to the symmetry of eq 1 in relation to E and $V(r)$, this approach can be easily implemented by simply redefining eq 1 as

$$\Delta V(r) = \begin{cases} \dfrac{(V(r) - E)^2}{\alpha + (V(r) - E)}, & V(r) \geq E \\ 0, & V(r) < E \end{cases} \quad (3)$$

In this implementation, the boost potential, $\Delta V(r)$, is subtracted from the true potential $V(r)$ whenever the potential

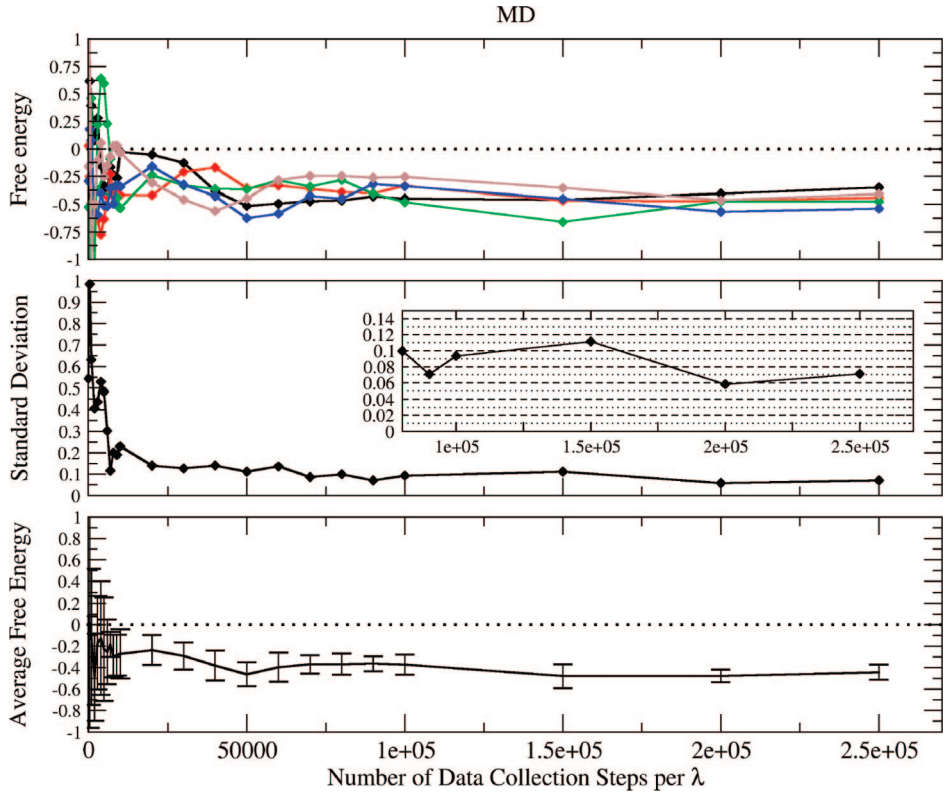

Figure 6. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

$V(r)$ is greater than the boost energy E; in this case, the simulation is performed on the modified potential $V^*(r) = V(r) - \Delta V(r)$. On the other hand, when $V(r)$ is below the energy level E, the simulation is performed on the true potential $V^*(r) = V(r)$. Hereafter, this approach will be referred to as aMD[b] and aMDt[b] when applied to the torsional terms of the potential energy. Figures 1 and 2 illustrate a schematic representation of a hypothetical one-dimensional potential modified using eqs 1 and 3, respectively. In both cases, as α decreases, the modified potential becomes flatter, and as α increases, the modified landscape asymptotically approaches the unmodified potential.

In this approach, analogously to Hamelberg et al.'s implementation, the correct canonical ensemble averages of an observable are fully recovered by reweighting each configuration by the Boltzmann factor of the negative of the boost potential energy, $\exp\{-\beta[\Delta V(r)]\}$. The application of this schema into the dual boost approach is straightforward. In this case, eq 2 is simply redefined as $V^*(r) = \{V_0(r)+[V_t(r)-\Delta V_t(r)]\}-\Delta V_T(r)$, and the boost factor as $\exp\{-\beta[\Delta V_t(r)+\Delta V_T(r)]\}$. This implementation will be referred to as aMDtT[b].

Coupling Accelerated MD Approach with Thermodynamic Integration Simulations. Thermodynamic integration is a commonly used technique to compute the difference in free energy between two thermodynamic states, which differ from each other according to their intermolecular or intramolecular interaction potentials.[3,8,12] In this case, the interaction potential can be expressed as a function of a coupling parameter, λ, that determines the state of the system.[10] Thus, by defining the free energy, F, as a continuous function of λ, the difference in free energy between two states is given by

$$\Delta F = \int_{\lambda=0}^{\lambda=1} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda \tag{4}$$

where $\lambda = 0$ and 1 correspond to the initial and final states, respectively. Since F(λ) can be written as

$$F(\lambda) = -k_b T \ln Q(\lambda) \tag{5}$$

ΔF can be rewritten as[45]

$$\Delta F = \int -\beta \frac{1}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} d\lambda \tag{6}$$

where Q is the partition function of the system, $\beta = 1/k_b T$, k_b is Boltzmann's constant, and T is the temperature. Here, we use the partition function for canonical ensemble, which is defined as[46]

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int dr dp \exp[-\beta H(p, r)] \tag{7}$$

where N is the number of particles, h is Planck's constant, p and r are the momenta and positions of the particles, and H is Hamiltonian of the system. Substituting eq 7 into eq 6 and deriving in respect to λ,[45] we obtain

$$\frac{\partial F(\lambda)}{\partial \lambda} = \frac{\int\int dp dr \frac{\partial \mathrm{H}(p, r)}{\partial \lambda} \exp[-\beta \mathrm{H}(p, r)]}{\int\int dp dr \exp[-\beta \mathrm{H}(p, r)]} \tag{8}$$

Assuming that the kinetic energy term is separable and not dependent on λ, eq 8 can be rewritten in terms of the potential energy $V(r)$ of the system

$$\frac{\partial F(\lambda)}{\partial \lambda} = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)]}{\int dr \exp[-\beta V(r)]} \tag{9}$$

and, finally

$$\Delta F = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{10}$$

where the integrand is the ensemble average of $\partial V/\partial \lambda$ calculated on the original potential $V(r)$ at a specific value of λ, and ΔF is the free energy difference between the initial ($\lambda = 0$) and final ($\lambda = 1$) states obtained on the unmodified potential surface, $V(r)$.

Similarly, for the modified potential we have

$$\frac{\partial F(\lambda)}{\partial \lambda}_* = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V^*(r)]}{\int dr \exp[-\beta V^*(r)]} \tag{11}$$

now, the ensemble average of true $\partial V(r)/\partial \lambda$ is performed over the modified potential $V^*(r)$.

Since both approaches can be coupled with TI simulations, we will first express $V^*(r)$ as

$$V^*(r) = V(r) + \Delta V \tag{12}$$

In this case, eq 11 can be rewritten as

$$\begin{aligned} \frac{\partial F(\lambda)}{\partial \lambda}_* &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp\{-\beta[V(r) + \Delta V]\}}{\int dr \exp\{-\beta[V(r) + \Delta V]\}} \\ &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]} \end{aligned} \tag{13}$$

The Boltzmann distribution can be extracted from the non-Boltzmann distribution using the method introduced by Torrie et al.[19] The corrected canonical distribution can then be recovered by reweighting the phase space of the modified potential by multiplying the integrand by the strength of the bias at each position, which in this case corresponds to $\exp[\beta \Delta V]$.

$$\begin{aligned} &\frac{\partial F(\lambda)}{\partial \lambda}_C \\ &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V]} = \frac{\partial F(\lambda)}{\partial \lambda} \end{aligned} \tag{14}$$

Thus, the corrected ensemble average of ΔF^C can be obtained by dividing both the numerator and the denominator of eq 11 by $\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] = \int dr \exp[-\beta V^*(r)]$

$$\begin{aligned} &\frac{\partial F(\lambda)}{\partial \lambda}_C \\ &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V] / \int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V] / \int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]} \end{aligned}$$

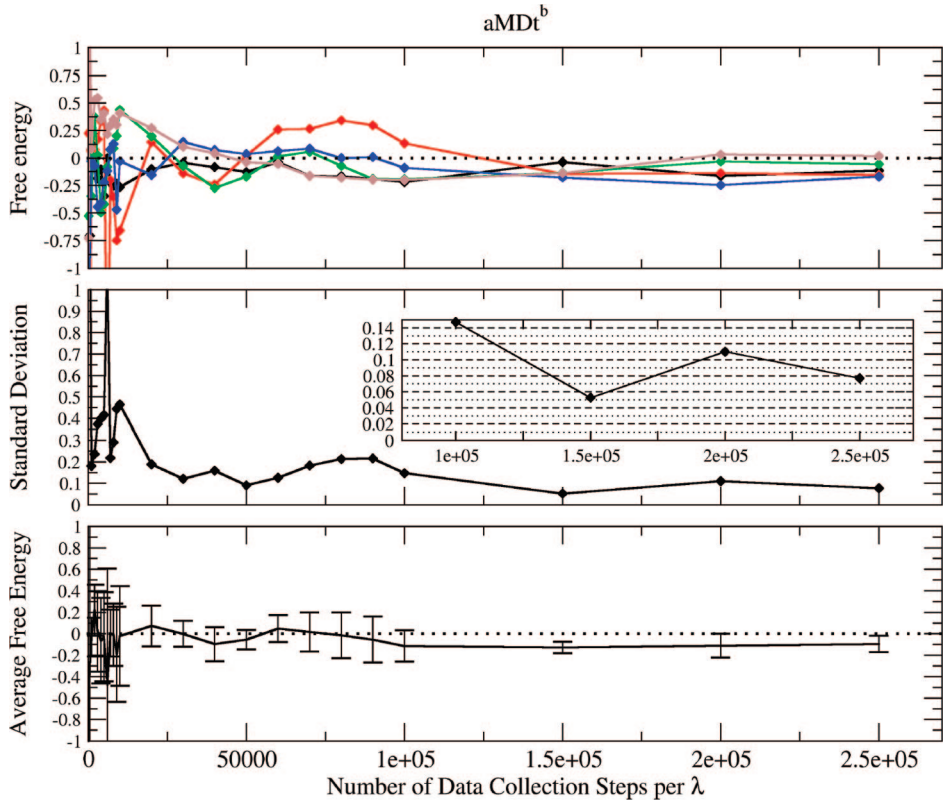

Figure 7. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 $\times$ 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

$$= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V^*(r)] \exp[\beta \Delta V] / \int dr \exp[-\beta V^*(r)]}{\int dr \exp[-\beta V^*(r)] \exp[\beta \Delta V] / \int dr \exp[-\beta V^*(r)]}$$

and integrating over λ.

$$\Delta F^C = \int_{\lambda=0}^{\lambda=1} \left[\left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \exp[\beta \Delta V] \right\rangle_{\lambda^*} / \langle \exp[\beta \Delta V] \rangle_{\lambda^*} \right] d\lambda$$

$$= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$$= \Delta F \tag{16}$$

Analogously, if the modified potential is defined as $V^*(r) = V(r) - \Delta V$, eq 16 can be redefined as

$$\Delta F^C = \int_{\lambda=0}^{\lambda=1} \left[\left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \exp[-\beta \Delta V] \right\rangle_{\lambda^*} / \langle \exp[-\beta \Delta V] \rangle_{\lambda^*} \right] d\lambda$$

$$= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$$= \Delta F \tag{17}$$

Therefore, independent of the approach applied, the accelerated molecular dynamics simulation method converges to the canonical distribution, and the corrected canonical ensemble average of the system is obtained by simply reweighting each point in the configuration phase space on the modified potential by the strength of the Boltzmann factor of the bias energy, $\exp[\beta \Delta V]$ or $\exp[-\beta \Delta V]$, at that particular point.

Results

The first question to be answered about the aMD^b approach is if this method is able to improve conformational transitions. To address this question, we performed MD simulations of a butane molecule in explicit water and monitored the dihedral angle shown in Figure 3. Figure 4 shows the result obtained from normal MD simulations. It is worth noting that even for a simple system like this, the number of conformational transitions is still very limited. Figure 4-middle displays the results obtained from the aMD^b approach. In this simulation, only the torsional term of the potential energy was applied in the boost potential (aMDt^b). Parameters E and α were set to 0.5 and 0.2 kcal/mol, respectively. For comparison, Figure 4-bottom shows the dihedral transitions calculated with the aMDt approach. In this case, parameters E and α were set to 5.0 and 0.5 kcal/mol, respectively. As expected, more conformational transitions are observed with aMDt^b and aMDt than with normal MD simulations. Figure 4 also reveals that, even though both methods improved conformational sampling, aMDt^b still produces a much larger number of transitions than aMDt.

Free Energy Calculations. The propane-to-propane system (Figure 5) was used to test convergence and accuracy of the accelerated TI simulations. A similar system, ethane-

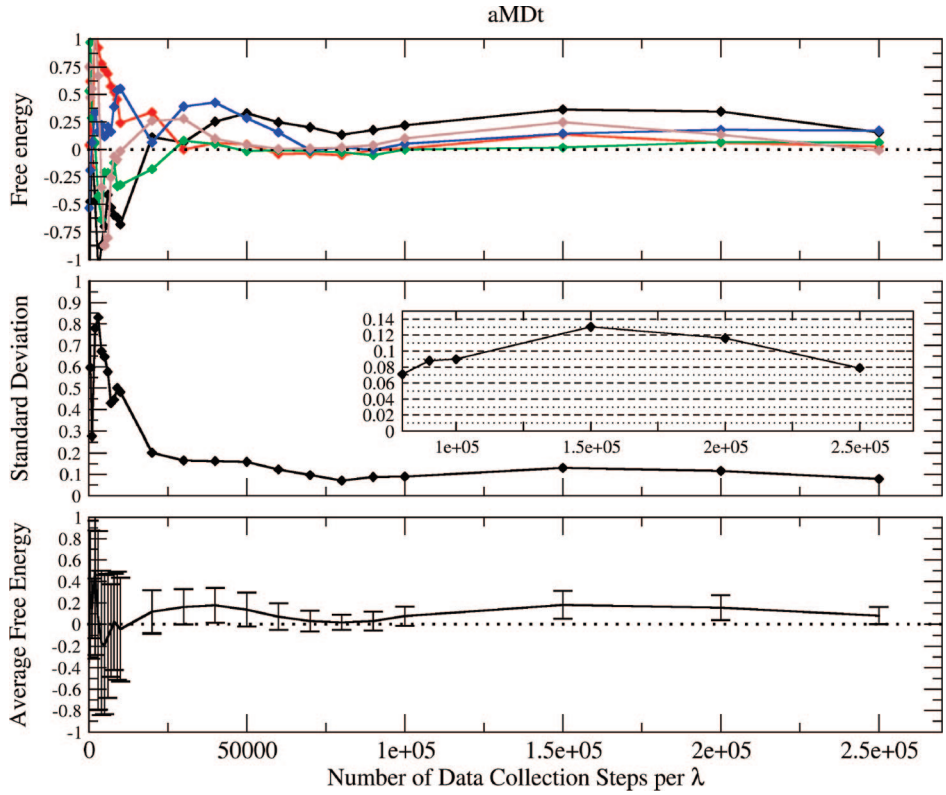

Figure 8. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 $\times$ 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

to-ethane transformation, has been used before by other groups to test the performance of different approaches to calculate free energy changes. This transformation is particularly interesting because independent of the force field, water model, or simulation method used, the free energy change should be equal to zero ($\Delta G = 0$).

The free energy changes obtained with normal MD simulations are compared to the ones obtained with aMDt^b, aMDt, and aMDTt^b. Figures 6, 7, 8, 9, and 10 show the free energy change, the average free energy, and the error associated with the propane-to-propane transformation calculated from five independent TI simulations. In all simulations, the same amount of sampling was performed at each window, and an equal amount of time was spent in equilibration and data collection. The error was estimated by calculating the standard deviation of the five independent simulations as a function of time. All TI calculations with normal MD fail to reproduce the expected free energy value (Figure 6), converging to free energy values of ≈ -0.4 kcal/mol. The change with time of the average free energy toward the expected free energy value is rather slow, and it is clear that this normal MD requires much longer simulations to reproduce accurate results. Figure 7 displays the TI results obtained with aMDt^b. Although the calculated average free energy change is closer to the corrected free energy value, like normal MD, longer simulations are still required to reproduce the correct average free energy change. Similar results were obtained when the aMDt approach was applied (Figure 8). However aMDt^b still seems to converge better than aMDt. Nevertheless, both approaches perform better than normal TI simulations, and this improvement can be mainly attributed to the increasing in conformational sampling.

As mentioned before, aMDt^b and aMDt approaches modify the energy landscape by adding a boost potential to the potential surface, and, in these cases, the boost potential is based on the torsional terms of the potential energy. Even though the conformational sampling is clearly enhanced in both approaches, those approaches still fail to generate accurate results. The reason for that might be the absence of energy terms in the boost potential describing solute−solvent and solvent−solvent interactions. Therefore, in order to also accelerate the solvent response along the propane-to-propane transformation, the dual boost approach was also tested with TI calculations (ΔV_T was applied with parameters E and α set to −3.0 and 30.0 kcal/mol per atom, respectively). Owing to instabilities introduced by the application of aMDTt approach in TI simulations, only results obtained with aMDTt^b are displayed in Figure 9. By comparing aMDTt^b and aMDt^b, we see clearly that inclusion of the potential energy terms describing solute−solvent and solvent−solvent interactions in eq 3 dramatically improves the accuracy and convergence of the TI simulations. It is worth mentioning that all calculated free energy values using aMDTt^b, with at least 100 ps of data collection, converged to the correct value and are within the estimated error. For the system studied in this work, aMDTt^b was the only approach to achieve

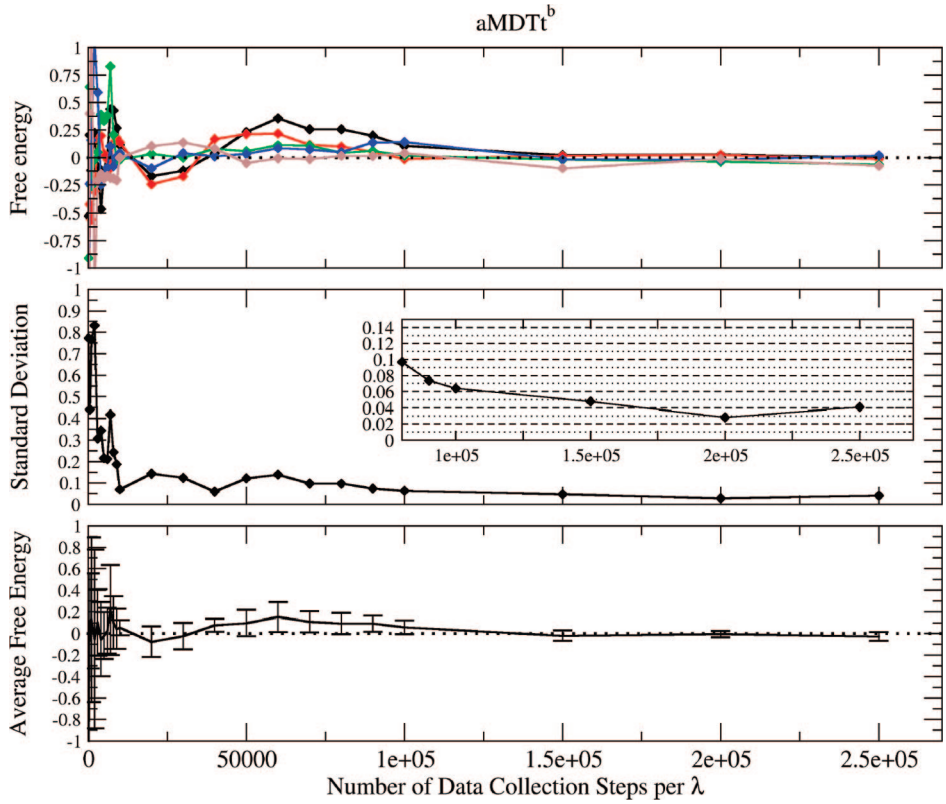

Figure 9. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 $\times$ 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

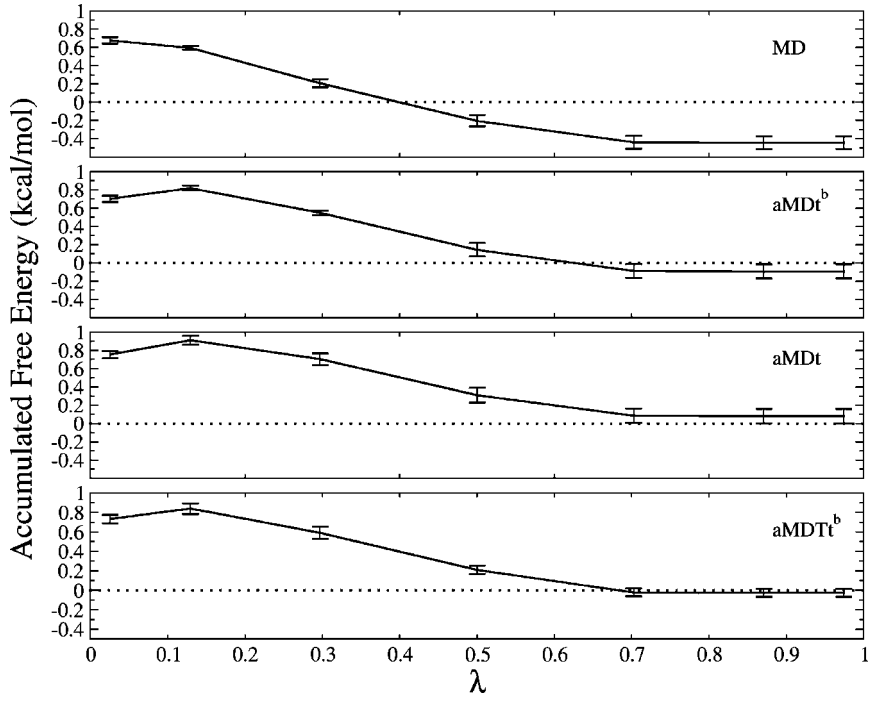

Figure 10. Accumulated free energy change for the propane-to-propane simulations calculated with 250 $\times$ 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ.

converged and accurate results from TI simulation (ΔG = -0.027 ± 0.04 kcal/mol).

Figure 10 shows the accumulated free energy for each value of λ. In this plot, the accumulated free energy was calculated by using equilibration = data collection time = 250 ps for each window. Except for the aMDTt^b, which performed remarkably well, all approaches failed to reproduce the corrected free energy change for the propane-to-propane transformation.

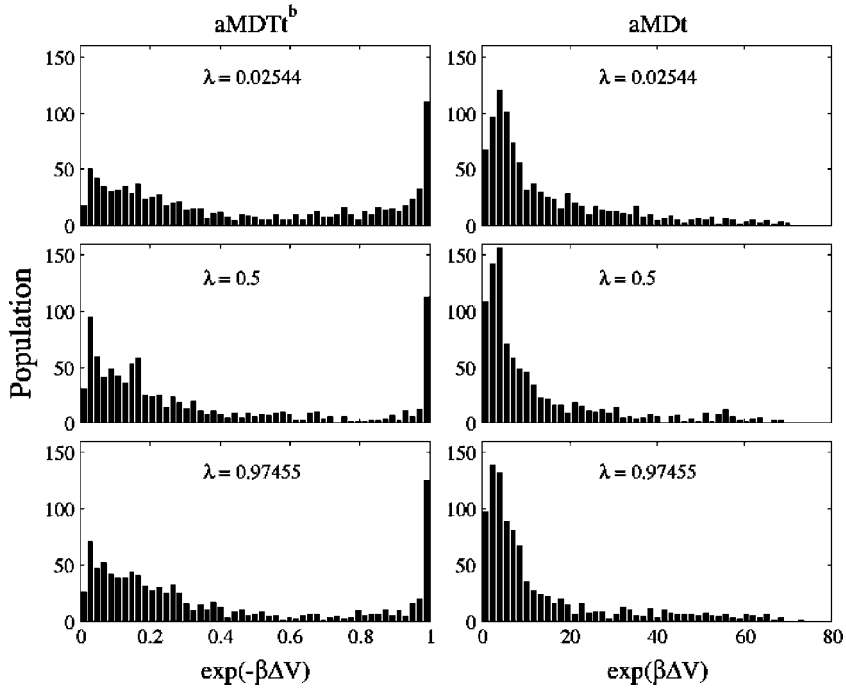

Figure 11. Distribution of Boltzmann factor of the boost potential calculated from propane-to-propane simulations using the aMDTt^b (left) and aMDt (right) approaches.

Distribution of the Boost Energy along the TI Simulations. The main difference between the two approaches (eqs 1 and 3) consists of how the modified potential surface is generated. For methods based on the aMD approach, the molecular motions are accelerated by raising the energy basins on the potential surface. Although this method proved to be excellent to enhance conformational sampling of biomolecules, some issues concerning the calculation of thermodynamics properties still need to be addressed. For instance, to fully recover ensemble average properties, each point in the phase space should be multiplied by its respective Boltzmann factor of the boost energy, $\exp(\beta\Delta V)$. In some cases, when this procedure is applied, relatively few configurations in the entire trajectory have significant contributions to the ensemble average. As a consequence, the statistics are compromised, and the thermodynamic property is not fully converged. This is the main issue to be addressed when the aMD method is coupled with TI calculations. It is worth mentioning that in this implementation ΔV is a nonnegative number, and its respective Boltzmann factor produces numbers in the interval $[1\rightarrow\infty)$. Thus, large values of $\exp(\beta\Delta V)$ correspond to configurations near energy minima of the potential surface, while small values correspond to relatively high-energy regions. Figure 11, on the right, displays the distribution of the boost factor along the aMDt simulations at three different values of λ. It is clear from those plots that, even for this rather low acceleration condition, the system spends almost the entire simulation in regions of relatively high energy. Only very few configurations (for instance configurations with $\exp(\beta\Delta V)>50$) will effectively contribute to the ensemble average.

In order to address this issue, here, we introduce the aMD^b approach aiming to improve sampling without compromising the statistics in the TI calculations. As mentioned before, in this approach, regions near the minima in the potential surface are left unchanged, and the Boltzmann factor of the boost energy is now defined as $\exp(-\beta\Delta V)$. In this implementation, $\exp(-\beta\Delta V)$ assumes values in the interval $(0\leftarrow 1]$. Thus, unlike the aMD approach, all configurations near the low-energy regions of the potential surface ($\Delta V = 0$), which are the ones that most contribute to the ensemble average, have the same weight of $\exp(-\beta\Delta V) \approx 1$. Besides that, configurations sampled in high-energy regions of the conformational space have rather small weights, $\exp(-\beta\Delta V) \ll 1$, and, as a consequence, have a fairly small contribution to the ensemble average. Figure 11, on the left, shows the boost factor distribution along the aMDTt^b simulations at three different values of λ. It is clear that the aMD^b approach is not only able to sample both low- and high-energy regions of the potential surface but also to keep regions near the minima well populated. As a consequence, the statistics are not compromised in the TI calculations, and the ensemble average is recovered (Figure 9).

Methods

All calculations were performed using the Sander module in the AMBER8[47] package that was modified to carry out the accelerated MD simulations. The GAFF force field was used to describe the solute in all simulations. The butane molecule was solvated in a periodic box of explicit TIP3P waters,[48] which extends on each side 10 Å from the closest atom of the solute, by using the Leap module in AMBER. To bring the system to its correct density, we carried out an MD simulation for 1 ns in which the NPT ensemble (T = 300 K, P = 1 atm) was applied. All data collection was carried out over MD simulations of 1 ns, during which the NVT ensemble (T = 300 K, density= 0.984 g/mL) was applied. The final configuration was then used as the starting point for the propane → propane simulations. In both

systems, butane and propane → propane simulations, each solute atom was assigned with zero partial charge. The free energy change was calculated by varying λ form 0 (initial state) to 1 (final state). All TI simulations were carried out using seven discrete points of λ, which were determined by Gaussian quadrature formulas. Normal and accelerated MD simulations of 500 ps were carried out for each λ point. The NVT ensemble was used in all TI simulations. Temperature and pressure were controlled via a weak coupling to external temperature and pressure baths[49] with coupling constants of 0.5 and 1.0 ps, respectively. Apart from all TI simulations where the time step was set to 1 fs, the equations of motion were integrated with a step length of 2.0 fs using the Verlet Leapfrog algorithm.[50] For further analysis, the trajectory was saved every 1.0 ps. The PME summation method was used to treat the long-range electrostatic interactions in the minimization and simulation steps.[51,52] The short-range nonbonded interactions were truncated using a 8 Å cutoff, and the nonbonded pair list was updated every 20 steps.

Conclusions

In this work, we showed a straightforward way of coupling the Thermodynamic Integration approach with the accelerated MD method. We also introduced a new approach, aMD^b, aiming to improve convergence and efficiency of free energy calculations in condensed-phase systems. The results obtained with aMD and aMD^b were compared with conventional TI calculations. Our results showed that both accelerated MD approaches improve conformation sampling when compared to normal MD simulations. When applied to just torsion terms of potential energy, both approaches, aMDt and aMDt^b, increased substantially the number of conformation transitions of the butane molecule in explicit water when compared to normal MD simulations. In addition, the accuracy of free energy simulations was significantly improved when sampling of internal degrees of freedom of solute was enhanced. However, accurate and converged results were only achieved when the solvent interactions were taken into account in the accelerated MD approaches. When combined with aMD^b, the application of dual-boost approach improved markedly the convergence and accuracy of TI calculations. By analyzing the distribution of the boost potential along the free energy simulations, we observed that the aMD^b approach efficiently samples both low- and high-energy regions of the potential surface. Since this approach also maintains well populated regions near the minima, the statistics are not compromised in the TI calculations, and, as a result, the ensemble average can be recovered.

Acknowledgment. This work was supported in part by grants from NSF, NIH, the Center for Theoretical Biological Physics, the National Biomedical Computation Resource, NSF Supercomputing Centers, and Accelrys, Inc.

References

(1) Tembe, B. L.; Mccammon, J. A. *Comput. Chem.* **1984**, *8*, 281.

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

(3) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175.

(4) VanGunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.

(5) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.

(6) Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(7) Mezei, M.; Beveridge, D. L. *Methods Enzymol.* **1986**, *127*, 21.

(8) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(9) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(10) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(11) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.

(12) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.

(13) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.

(14) Straatsma, T. P.; McCammon, J. A. *Methods Enzymol.* **1991**, *202*, 497.

(15) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589.

(16) Jorgensen, W. L.; TiradoRives, J. *J. Phys. Chem.* **1996**, *100*, 14508.

(17) Min, D. H.; Li, H. Z.; Li, G. H.; Bitetti-Putzer, R.; Yang, W. *J. Chem. Phys.* **2007**, *126*, 144109.

(18) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93*, 2769.

(19) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(20) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865.

(21) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.

(22) Itoh, S. G.; Okamoto, Y. *J. Chem. Phys.* **2006**, *124*, 104103.

(23) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.

(24) Grubmuller, H. *Phys. Rev. E* **1995**, *52*, 2893.

(25) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222.

(26) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339.

(27) Darve, E.; Wilson, M. A.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113.

(28) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(29) Kim, J.; Straub, J. E.; Keyes, T. *J. Chem. Phys.* **2007**, *126*, 135101.

(30) Li, H.; Min, D.; Liu, Y.; Yang, W. *J. Chem. Phys.* **2007**, *127*, 094101.

(31) Ceotto, M.; Ayton, G. S.; Voth, G. A. *J. Chem. Theory Comput.* **2008**, *4*, 560.

(32) Xing, C. Y.; Andricioaei, I. *J. Chem. Phys.* **2006**, *124*, 034110.

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

(34) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(35) Voter, A. F. *J. Chem. Phys.* **1997**, *106*, 4665.

(36) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 155102.

(37) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 13778.

(38) Hamelberg, D.; Mongan, J.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 76.

(39) Hamelberg, D.; Shen, T.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 1969.

(40) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724.

(41) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532.

(42) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487.

(43) Hamelberg, D.; Shen, T.; McCammon, J. A. *J. Chem. Phys.* **2005**, *122*, 241103.

(44) Hamelberg, D.; Shen, T. Y.; McCammon, A. *Biophys. J.* **2005**, *88*, 183A.

(45) Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R. *J. Chem. Phys.* **1983**, *79*, 2375.

(46) Mcquarrie, D. A. *Phys. Today* **1965**, *18*, 74.

(47) Case, D. A.; Perlman, D. A.; Caldwell, J. W.; Chetham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. 2002.

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

(49) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(50) Hockney, R. W. *Bull. Am. Phys. Soc.* **1968**, *13*, 1747.

(51) Ding, H. Q.; Karasawa, N.; Goddard, W. A. *J. Chem. Phys.* **1992**, *97*, 4309.

(52) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

CT800160Q