Acc. Chem. Res. **2006,** *39,* 73−81

ARTICLES

Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions

BERND ENSING,*,† MARCO DE VIVO,† ZHIWEI LIU,‡ PRESTON MOORE,‡ AND MICHAEL L. KLEIN†

Center for Molecular Modeling and Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104-6323, and Department of Chemistry & Biochemistry, University of the Sciences in Philadelphia (USP), 600 South 43rd Street, Philadelphia, Pennsylvania 19104-4495

Received May 11, 2005

ABSTRACT

The metadynamics or hills method is a relatively new molecular dynamics technique aimed to enhance the sampling of separated regions in phase space and map out the underlying free energy landscape as a function of a small number of order parameters or collective variables. The high efficiency allows for the application of metadynamics in combination with first principles dynamics methods, in particular with Car−Parrinello molecular dynamics, to study processes in which changes in the electronic structure play a dominant role, such as chemical reactions. The option to choose several independent collective variables is important to tackle complex and concerted transformations that lack an obvious a priori choice for a single reaction coordinate. In this Account, we discuss the role of metadynamics in the search of transition states, local minima, reaction paths, free energy profiles, and reaction coordinates among a growing list of alternative methods.

I. Introduction

Free energy simulations have been providing an increased understanding of the driving forces involved in a wide range of interesting phenomena in chemistry and biophysics.[1,2] Knowledge of the free energy landscape, the stable basins, the separating saddle points, and the connecting reaction paths is essential to learn the direction of the process, the relative stability of intermediate and final states, and the reaction mechanism of the process. The difference between the free energy of the transition state and that of the reactants is used to obtain the reaction rate within the transition state theory approximation, which is an upper bound of the true rate.

For systems consisting of only several tens of atoms in vacuo, optimization of the reactants and products on the 0 K potential energy surface (PES) is routinely performed, adding temperature and entropy effects a posteriori within the harmonic approximation. The transition state and connecting minimum energy path can be localized using the local curvature of the PES via quasi-Newton methods or other advanced techniques (see ref 3 and references therein). However, many complex phenomena such as chemistry in liquids, protein folding, and phase transitions require direct sampling of phase space because the optimized structure does not resemble the configurational distribution at finite temperature (for example, the solvent structure of solutes in water is not well modeled by the solutes in optimized water, being crystalline ice), whereas in practice the optimization followed by the frequency calculation is very cumbersome or even impossible.

Free energy methods such as free energy perturbation[4] (FEP) are very powerful in calculating, for instance, the difference in hydration free energy between two solutes or the difference in binding free energy between ligands. Typically, FEP proceeds by stepwise transmutation of one of the species into the other (therefore often referred to as computational alchemy) via some artificial path based on a superposition of the two species. Here however, we will focus our attention on the application of free energy methods to enhance the sampling of interesting regions in phase space that are either so high in (free) energy that their sampling during a standard molecular dynamics (MD) simulation would be a rare event (e.g., transition states) or separated from the initial basin by such high energy passes (namely, the product states).

Traditional free energy methods for this purpose can be categorized into two groups, namely, (1) methods based on umbrella sampling,[1] which apply an artificial biasing potential (such as the negative of the underlying free energy profile[5,6] or overlapping parabolic window potentials[7]) to enhance sampling of unfavorable regions along a reaction coordinate and (2) methods based on thermodynamic integration[8] in which case the so-called potential of mean force (i.e., the free energy function) is obtained from measurements of the thermodynamic force working on a reaction coordinate. Constrained MD[1] and adaptive biasing force MD[9] belong to the latter group, as does the popular steered MD[10] method, which is based

Dr. Bernd Ensing (http://www.cmm.upenn.edu/∼ensing) is currently transitioning from a postdoctoral fellowship with Prof. Klein to a second fellowship with Prof. Parrinello.

Dr. Marco De Vivo (http://www.cmm.upenn.edu/∼mdevivo) is currently a postdoctoral fellow with Prof. Klein.

Dr. Zhiwei Liu (http://hydrogen.usip.edu/zliu) is currently a staff researcher at USP.

Prof. Preston B. Moore (http://hydrogen.usip.edu/moore) is currently an associate professor at USP.

Prof. Michael L. Klein (http://www.lrsm.upenn.edu/) is Hepburn Professor of Physical Science and Director of the LRSM at UPENN.

* To whom correspondence should be addressed. E-mail: ensing@cmm.upenn.edu.

† University of Pennsylvania.

‡ University of the Sciences in Philadelphia (USP).

10.1021/ar040198i CCC: $33.50
Published on Web 11/23/2005

on nonequilibrium simulations using Jarzynski's equality.[11] Both types of methods heavily rely on a good choice for a reaction coordinate, that is an order parameter that describes a reversible path from reactants to products. In these cases, choosing a physically appealing path, for example, one based on the unstable normal mode in the transition state, helps to avoid hysteresis problems.

Finding a good reaction coordinate poses one of the most challenging problems in many applications of free energy calculations, especially when they are concerned with complex collective transformations, as is the case with solvent rearrangements during a chemical reaction in solution, protein folding, or the efficient concerted biochemical reactions catalyzed in the active sites of enzymes. A special group of methods focuses on finding a physical path that is not only indicative for the reaction mechanism but can also be used as the reaction coordinate in traditional free energy methods. The dimer method and the nudge elastic band method are examples of such methods that trace the minimum energy path in the 0 K PES (see ref 3 and references therein), whereas other methods also include temperature and entropic contributions, for example, by minimizing an action integral.[12,13] Moreover, these paths can be used as the starting point for the generation of new reactive pathways, using the transition path sampling[14] (TPS) method, which aims to naturally sample the ensemble of all possible reactive paths via a Metropolis criterion. Although computationally demanding, the power of TPS lies in the facts that it captures the true reactive dynamics and that it only requires collective variables that specify the stable states instead of a connecting reaction coordinate.

In recent years, a number of novel advances have been developed featuring alternative approaches to enhance the sampling of rare events, which can again be subdivided into two main groups. The first group of methods uses elevated temperatures to overcome barriers. For example, by simulating simultaneously several replicas of the system at different temperatures and allowing swap moves between replicas (known as parallel tempering[15]), the system of interest, which is the replica with the lowest temperature, "jumps" over barriers as configurations from the replica at the highest temperature trickle down along the chain of replicas. Alternatively, excitation of a small number of order parameters or collective variables (e.g., internal bond distances) is achieved by raising their kinetic energy within a single simulation to overcome barriers.[16−18]

The second group of novel methods builds on umbrella sampling (US). Already in the early days, it was recognized that the hardest part of US is the construction of a good biasing potential that counterbalances the intrinsic free energy barrier,[6] which led to the development of adaptive biasing potential methods that improve the potential on the fly.[5,19,20] A special subclass of these history-dependent biasing potential methods builds a repulsive potential in the low-energy basins rather than constructing an attractive potential in the TS region. Examples are the local elevation method,[21] conformational flooding,[22] and the Monte Carlo method by Wang and Landau.[23]

The metadynamics method,[24] which is the focus of this Account, falls in this last subcategory. In brief, the algorithm places in a coarse time line repulsive markers in the space spanned by a small number of chemically relevant collective variables. These markers or little "hills" placed on top of the underlying free energy landscape (hence the pseudonym "hills method") discourage the system from revisiting points in configurational space and rapidly accumulate in the initial basin, allowing the system to escape over the lowest transition state as soon as the growing biasing potential counterbalances the underlying free energy well. This way, the method not only accelerates simulation of rare events by effectively escaping free energy minima and exploring new pathways but also maps out the underlying free energy surface as the negative of the sum of potential hills. Metadynamics is not designed to find the perfect reaction coordinate. However, the possibility to, in principle, keep expanding the set of collective variables (to the point where the method becomes impractical) increases the probability of capturing all the relevant barriers, which is verified by monitoring potential hysteresis similar as done with US.

Metadynamics has been applied in several fields with quite some success, including applications in biophysics, chemistry, material science, and crystal structure prediction.[25] Also, the method has undergone a number of improvements and extensions since its original introduction by Laio and Parrinello.[24] For example, in ref 29 the evolution of the collective variables was changed from a discrete algorithm into a continuous one, in conjunction with molecular dynamics schemes. Also the size of the hills was made into a history-dependent adaptive factor that adapts to the underlying free energy surface.[29,33] Wu et al. introduced a version of metadynamics that is particularly efficient in mapping out the 0 K potential energy surface.[27] Another important extension is an algorithm that localizes the lowest free energy path (LFEP) connecting the reactant and product states and that integrates the free energy surface with respect to the directions perpendicular to the LFEP, mapping a multidimensional problem back to an intuitive one-dimensional free energy profile along a parametrized reaction coordinate.[34] And finally, a valuable development is the growing list of (exotic) types of collective variables (e.g., coordination numbers,[29] the number of hydrogen bonds,[26] lattice parameters,[31] relative molecule orientation/rotation,[28] and the potential energy[32]), which continuously broadens the scope of possible applications.

To give the reader a flavor of the possibilities offered by the metadynamics method, we will start with a brief introduction of the algorithm, followed by a discussion of three paradigmatic applications that illustrate the potential of this approach to overcome the rare event problem in complex systems.

II. Method

The metadynamics method aims to enhance the exploration of the free energy surface, $F(\mathbf{S}(\mathbf{R}))$ of a limited set of

collective variables $\mathbf{S}(\mathbf{R})$ (bond lengths, torsion angles, coordination numbers, etc.). For a canonical ensemble, the fundamental thermodynamic function is the Helmholtz free energy,

$$F = -k_B T \ln[Q(N,V,T)] \tag{1}$$

with T the absolute temperature, k_B Boltzmann's constant, and $Q(N,V,T)$ the canonical partition function. The probability to find the system on a hyperplane, $\mathbf{S}'$ of atomic positions, $\mathbf{R}$, in phase space is given by

$$p(\mathbf{S}') = \frac{1}{Q}\int d\mathbf{R}\, d\dot{\mathbf{R}}\, e^{(-H(\mathbf{R},\dot{\mathbf{R}})/k_B T)}\, \delta(\mathbf{S}(\mathbf{R}) - \mathbf{S}') \tag{2}$$

where $\dot{\mathbf{R}}$ are the atomic velocities, H is the Hamiltonian of the system, and δ is the delta function.

Molecular dynamics (MD) simulation is used to sample phase space, where we can derive a model Hamiltonian of the molecular system from the Lagrangian,

$$L_{MD} = \frac{1}{2}\sum_I M_I \dot{\mathbf{R}}_I^2 - V(\mathbf{R}) \tag{3}$$

A typical model potential, or force field, consists of sums of pairwise potentials for the atomic bond stretching, bending, torsional rotation, out-of-plane bending, intramolecular repulsion, van der Waals attraction, and electrostatic interaction as follows:[1]

$$V_{MM}(\mathbf{R}) = \frac{1}{2}\sum_{N_b} k_b (b - b_0)^2 + \frac{1}{2}\sum_{N_a} k_a (a - a_0)^2 + \sum_{N_t} k_t [1 + \cos(mt - t_0)] + \frac{1}{2}\sum_{N_o} k_o (o - o_0)^2 + \sum_I \sum_{J>I} \left[\frac{C_{12,IJ}}{R_{IJ}^{12}} - \frac{C_{6,IJ}}{R_{IJ}^{6}} + \frac{q_I q_J}{R_{IJ}} \right] \tag{4}$$

This so-called molecular mechanics (MM) force field is computationally efficient and allows for simulation of molecular systems of the size of hundreds of thousands of atoms for tens of nanoseconds on fast parallel computers. The third application presented hereafter (computation of the Ramachandran plot of the alanine dipeptide) uses the well-known CHARMM[35] MM force field.

Despite the usefulness of MM force fields, their functionality and parametrizations in general are based on approximations, such as harmonic bond stretching and the neglect of polarization and other many-body terms, that hinder or even exclude their applicability for certain interesting systems. Examples are the metallic centers of enzymes, systems with highly charged counterions, and most importantly, systems in which chemical bonds are broken and formed. Such complicated phenomena are modeled well when the interatomic forces are computed from the instantaneous electronic structure via an interface with an ab initio method. Density functional theory (DFT) is a particularly efficient electronic structure method, where the electronic (Kohn−Sham) energy, which is a functional of the electronic density $E_{KS} = E[\rho]$, is constructed from the one-electron wave functions ($\rho(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$).[36] The DFT potential,

$$V_{DFT}(\mathbf{R},\mathbf{r}) = \sum_I \sum_{J>I} \frac{q_I q_J}{R_{IJ}} - \frac{1}{2}\sum_i \int d\mathbf{r}\, \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) + \int d\mathbf{r}\, V_N(\mathbf{R})\rho(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}\, d\mathbf{r}'\, \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|r_{ij}|} + E_{XC}[\rho(\mathbf{r})] \tag{5}$$

contains terms accounting for the electronic kinetic energy, the electron−nuclei interaction, the Coulombic potential, and the so-called exchange−correlation potential. Given a configuration of atomic positions R, the meaningful (ground-state) potential is obtained from those electronic wave functions ψ that minimize V_{DFT}, which are readily found from a self-consistent matrix diagonalization.

Of course, during an MD simulation the atomic positions continuously change so that at every MD step the electronic wave functions have to be re-optimized to keep the electrons in the ground-state. Clearly, this is the most expensive part of the calculation. However, rather than performing the diagonalization at every step, the wave functions can be updated using an extended Lagrangian technique known as the Car−Parrinello MD (CPMD) method,[37]

$$L_{CPMD} = L_{MD} + \frac{1}{2}\sum_i \mu_i \int d\mathbf{r}\, |\dot{\psi}_i|^2 - V_{DFT}(\mathbf{R},\mathbf{r}) + \sum_i \sum_j \Lambda_{ij} \left(\int \psi_i^* \psi_j\, d\mathbf{r} - \delta_{ij} \right) \tag{6}$$

That is, the wave functions are treated as fictitious particles with a mass, μ_i, that follow the nuclei adiabatically for small enough μ_i as the potential V_{DFT} (instead of V_{MM}) evolves. The last term in eq 6 ensures wave function orthogonality through the Lagrange multipliers, Λ_{ij}.

In section B, we will use a hybrid, so-called quantum mechanics/molecular mechanics (QM/MM), method (see ref 38 and refs therein). In that case, the system is partitioned into a chemically active part, treated with QM, and the rest, treated with a MM force field via a mixed Lagrangian,

$$L_{QM/MM} = L_{CPMD} + L_{MM} - \sum_{i \in MM} q_i \int d\mathbf{R}\, \rho(\mathbf{r}) V_i(|\mathbf{R} - \mathbf{R}_i|) \tag{7}$$

with the last term accounting for the electrostatic interaction between the MM and QM parts of the system.

Having introduced the model Lagrangians for classical MD and first-principles CPMD, we will now focus our attention on enhancing the sampling of the collective variables $\mathbf{S}(\mathbf{R})$ that describe the (chemical) transition. Equation 2 shows that the probability to observe the system in the transition state during a simulation decreases exponentially with height of the TS barrier; a chemical TS barrier is typically one or two orders higher than the average kinetic energy ($k_B T \approx 0.6$ kcal/mol). The

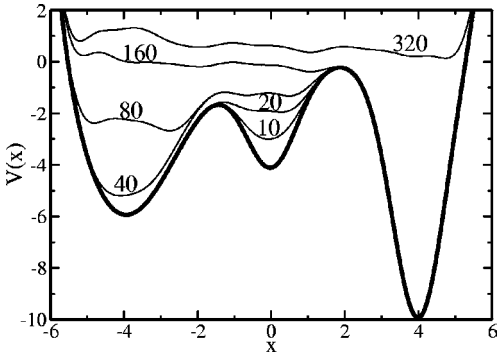

FIGURE 1. Filling a 1D model potential with hills, starting in the middle well. After 20 hills, the system escapes to the well on the left and so forth. Reproduced with permission from ref 24. Copyright 2002 National Academy of Sciences.

metadynamics method increases this probability by lowering the probability of revisiting phase-space regions S' in the low-energy reactant and product states in the following manner:

Admitting our soft spot for extended Lagrangian methods, we introduce a fictitious particle, s_α, for each collective variable, $S(\mathbf{R})$, with a mass M_α, which interacts with the system via a harmonic spring with force constant k_α attached to $S_\alpha(\mathbf{R})$,

$$L_{\text{meta}} = L_{\text{(CP)MD}} + \frac{1}{2}\sum_\alpha M_\alpha \dot{s}_\alpha{}^2 - \frac{1}{2}\sum_\alpha k_\alpha (s_\alpha - S_\alpha(\mathbf{R}))^2 + V_{\text{bias}}(\mathbf{s},t) \quad (8)$$

With a relatively large mass and stiff spring constant, these fictitious particles slowly "roll" in the bottom of the initial free energy well. However, every time interval, δt, we add a relatively small Gaussian-shaped repulsive potential

$$V_{\text{bias}}(\mathbf{s},t) = \sum_{t_i} H \exp\left(-\frac{|\mathbf{s} - \mathbf{s}(t_i)|^2}{2w^2}\right) \quad (9)$$

at the current point $\mathbf{s}(t)$ to the biasing potential $V_{\text{bias}}(\mathbf{s},t)$, which discourages the system from revisiting this point. The history-dependent potential builds up until it counterbalances the underlying free energy well, allowing the system to escape via a saddle point to a nearby local minimum, where the procedure is repeated. When all minima are "filled" with Gaussian potential "hills", the system moves barrier-free among the different states (Figure 1). The free energy is obtained as $-V_{\text{bias}}(\mathbf{s},t)$ to arbitrary resolution, depending on hill size (H,w) and time interval (δt).

III. Applications

We apply the metadynamics method to three illustrative problems that demonstrate the applicability and performance of metadynamics to efficiently explore and map out multidimensional free energy landscapes.

A. Assessment of Metadynamics. As pointed out in the Introduction, there is a growing list of methods that specialize in computing free energy profiles, accelerating rare events, and finding reaction pathways. Depending on which of these properties one wishes to compute, the specific application, and the computational resources available, a choice has to be made based on the performance and limitations of the available methods. For the metadynamics method, a quantitative assessment of the accuracy and efficiency in computing these properties was performed using Langevin dynamics, classical MD, and Car−Parrinello MD.[33,34,39]

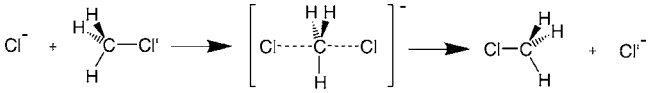

FIGURE 2. S_N2 reaction between Cl^- and CH_3Cl showing the symmetric transition state and the CH_3 conversion of configuration known as Walden inversion.

Laio et al applied metadynamics within Langevin dynamics to known free energy surfaces (FES) and derived explicit expressions for the error, ϵ (i.e., the deviation of the biasing potential from the FES), as a function of the parameters of the algorithm (H, w, δt) and system specific parameters, such as the system size, S, spanned by the collective variables, the temperature of the system, T, and the diffusion coefficient, D,

$$\epsilon = C(d)\sqrt{\frac{HTSw}{D\delta t}} \quad (10)$$

where $C(d)$ is a constant that depends on the dimensionality d (i.e., the number of collective variables).[39] Using a realistic case within classical MD, namely, the bidimensional free energy of a cyclophane-catenane complex in solution, they showed that this expression predicts the error with remarkable precision.[39]

Within the computationally demanding Car−Parrinello methodology, metadynamics parametrization for optimal efficiency is particularly important. Also the adiabatic decoupling of the collective variable dynamics from the auxiliary electronic wave function dynamics requires additional attention to avoid deviation from the Born−Oppenheimer ground state. Using the well-tested fundamental S_N2 reaction between Cl^- and CH_3Cl (shown in Figure 2), the performance of metadynamics was contrasted to standard umbrella sampling (US) calculations.

To this end, we developed an algorithm that localizes the lowest free energy path (LFEP) connecting the stable reactant and product states in the free energy surface resulting from a metadynamics simulation.[34] This LFEP is indicative for the reaction mechanism, in a way similar to the intrinsic reaction coordinate (IRC), but also includes temperature effects. Moreover, it maps the multidimensional landscape back to an intuitive one-dimensional parameter that is in fact a very good reaction coordinate to be used in traditional free energy methods such as US. Integration of the FES over the directions perpendicular to this reaction coordinate using a Monte Carlo US simulation results in the one-dimensional reaction free energy profile, which can be used as the biasing potential in an US simulation.

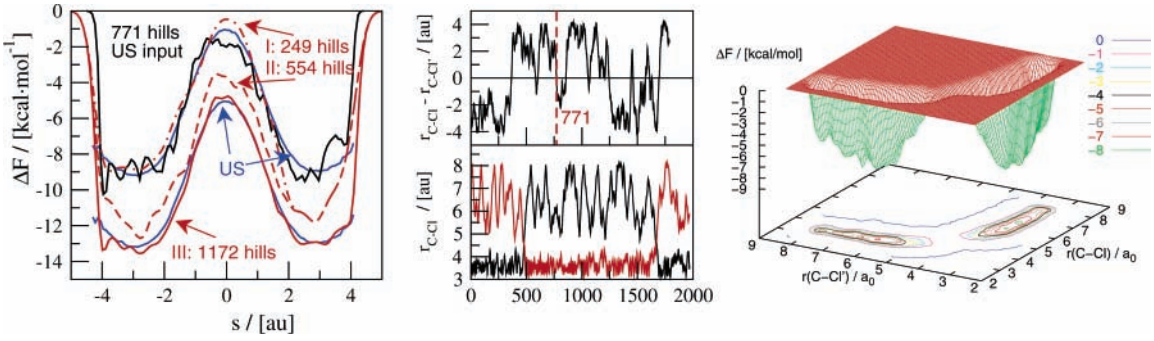

FIGURE 3. Left, free energy profiles resulting from metadynamics using the "difference of C−Cl distances" collective variable (black and red lines) compared to umbrella sampling (blue line); right, free energy surface using two collective variables, namely, the two C−Cl distances; center, metadynamics of the variables versus the number of hills added.[34]

Figure 3 shows the metadynamics results for the S_N2 reaction, once using a single collective variable (namely, the difference of the C−Cl distances) and once using two collective variables (i.e., the two C−Cl distances independently). The dynamics of these collective variables is shown in the center panel in Figure 3. The black line atop in this panel shows that the system escapes the reactant well, centered around $S(\mathbf{R}) \approx -3$ au, after adding 363 Gaussian hills and recrosses the TS after a total of 771 hills. The negative of the sum of these 771 hills is presented in the left panel (black line) and used (after smoothing) as the biasing potential in US. This initial metadynamics result after visiting and filling the reactant and product states only once compares already qualitatively favorably with the accurate US result (blue line), although quantitatively the profile is rather bumpy. Similarly, the bidimensional free energy surface of the two independent C−Cl distances was computed (shown in the right panel of Figure 3, together with the dynamics of the two collective variables in the center panel). Also here the potential along the LFEP was used to converge the one-dimensional free energy profile by US. This simple example showed that metadynamics can replace the often difficult problem of finding a single reaction coordinate with choosing relatively simple collective variables, such as the distances of the broken and formed chemical bonds.

Using the US result as a reference, we were able to test and fine-tune the metadynamics, which provided a range of parametrizations for which metadynamics performs efficiently and accurately.[34] The red lines in the left panel of Figure 3 illustrate the result of such a parametrization, where we reduced the hill size after every TS recrossing leading to efficient free energy convergence in three stages and a total of 1172 hills. Moreover, we were able to present a working recipe for obtaining the reaction mechanism and free energy profile for intrinsically multidimensional reactions, which was reported in ref 34.

Another example of the power of this recipe was its application to the prototypical base-induced bimolecular elimination (E2) reaction between F^- and CH_3CH_2F to form C_2H_4, HF, and F^-, for which, in contrast to the previous S_N2 reaction, there is no a priori choice for a simple reaction coordinate (see also Figure 4). Moreover, the free energy landscape shows additional product wells

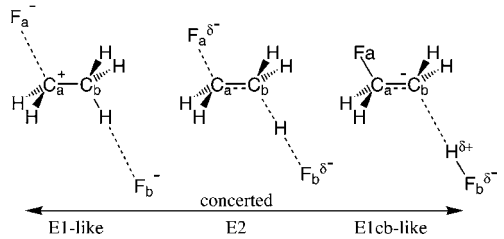

FIGURE 4. Mechanistic spectrum of possible transition states for the E2 reaction: left, the leaving group F^- leaves first; right, the proton leaves first; center, synchronous F^- and HF elimination.[30]

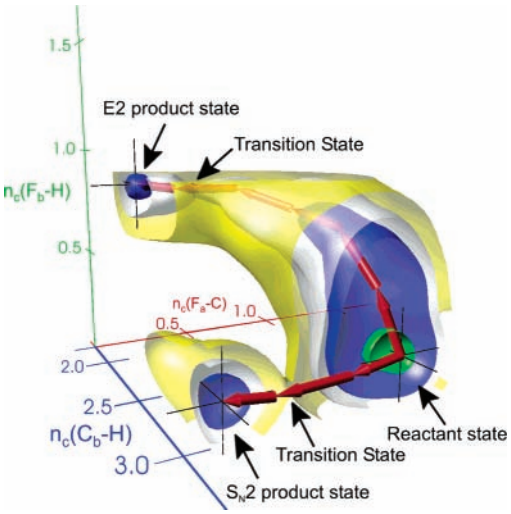

FIGURE 5. Three-dimensional projection of the 6D free energy surface showing the stable states (black crosses) and the E2 and S_N2 reaction channels. The red arrows depict the lowest free energy paths. Reproduced with permission from ref 30. Copyright 2005 National Academy of Sicences.

and reaction channels from competing S_N2 reactions (because the attacking F^- can attack the C_a). Using the metadynamics method with six collective variables, we were able to explore the stable states and connecting bottlenecks, after which we traced the LFEPs through the reaction channels and refined the free energy profiles for the E2 and S_N2 reactions using umbrella sampling.[30] Figure 5 shows a 3D projection of a relevant part of the 6D FES constructed from hills.

B. Solvent Effects on the S_N2 Reaction in Water. A common problem in the application of traditional free energy methods that depend on a good choice for a

reaction coordinate, such as constrained MD, steered MD, and umbrella sampling, to complex systems, is that often a posteriori sampling problems indicate that other slow degrees of freedom should have been included in the reaction coordinate. A classic example is the S_N2 reaction between Cl^- and CH_3Cl in aqueous solution, which has been studied both classically with US[40] and with constrained CPMD.[41] Both studies used a reaction coordinate based on the two C−Cl distances, thereby neglecting the slow change in the water coordination by the chloride ion during the reaction. Initially the coordination shell of the attacking Cl^- contains five H_2O molecules, which reduces to three in the TS and further to one in the product state, and vice versa for the leaving Cl.

In this second illustrative application, we show that metadynamics can be used to improve the sampling in such cases by systematically expanding the set of collective variables. Also, we will make use of a number of simple tricks that allows for the employment of metadynamics to complex and otherwise computationally demanding problems with much less effort. First of all, we will make use of the multiscale QM/MM approach as implemented in CPMD, which allows us to restrict the computational demanding QM part to only the chemically reactive S_N2 part of the system, while treating the rest of the system (the solvent) at the classical MM level of theory.

Second, we first reproduce the above-mentioned sampling problem by using only a single reaction coordinate in a constrained MD simulation and then employing metadynamics to repair the sampling problem. Obviously, we prefer to make use of the already obtained information of the FES, rather than starting again from scratch with filling the wells with hills. To this end, we reproduced the constrained CPMD calculation[41] in the QM/MM representation, using the difference of the two C−Cl distances as the reaction coordinate. The resulting potential of mean force obtained from 12 constrained MD runs with reaction coordinate values between −4.0 and 0.0 a_0 is in good agreement with the previous work. The free energy of the TS with respect to the bottom of the reactant well is 22.7 kcal/mol, which is 1 kcal/mol higher than that found for this S_N2 reaction in a periodic box of 32 QM treated water molecules. The sampling problem due to neglect of the slow solvent rearrangement was roughly estimated to lead to an overestimation of the barrier by 3 kcal/mol.[41]

We want to use this one-dimensional profile as the starting biasing potential in a metadynamics simulation. That is, to include the neglected solvent effects, we want to perform a 3D metadynamics simulation using the same reaction coordinate plus two collective variables that describe the changing coordination of the two chlorides. The latter description is functionalized as

$$n_c(\mathrm{Cl{-}H}) = \sum_{i=1}^{N_H} \frac{1-\left(\dfrac{r_{\mathrm{Cl-H}_i}}{r_0}\right)^8}{1-\left(\dfrac{r_{\mathrm{Cl-H}_i}}{r_0}\right)^{16}} \qquad (11)$$

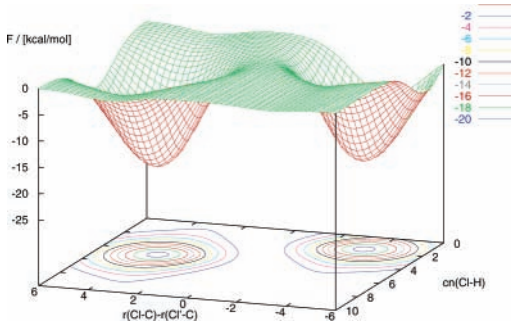

FIGURE 6. Potential of mean force of the S_N2 reaction in water obtained from constrained QM/MM CPMD.

which estimates the Cl−H coordination number by counting the number of solvent hydrogens within a radius $r_0 = 4.5\ a_0$ of the chlorine.

To be able to make use of the potential of mean force already obtained from the constrained MD, it has to be transformed into a 3D biasing potential. The two unknown dimensions are approximated by histogramming the two coordination numbers, n_c(Cl−H) during each of the constrained MD simulations and transforming these normal probabilities $P(n_c)$ into potentials of mean force, $V(n_c) = k_B T \ln[P(n_c)]$, which were fitted by parabola. These potentials were used to construct a 3D FES, the reactant and product wells of which were fitted with a sum of Gaussian hills. The resulting potential, shown in Figure 6, was then used as the starting bias for the metadynamics simulation.

The final technique for further speed-up exploits the symmetry that exists between the reactant and product states, by adding simultaneously additional hills in the symmetrically equivalent points in the collective variables space every time that a hill is added to the history-dependent potential. This way, the product well is filled automatically, while the reactant well is being filled without actually having to visit the product state, or vice versa. Note that the symmetry feature can also be used when the symmetry is not perfect (e.g., if one of the two Cl's were a Br) to gain an initial speed-up to fill most of the wells, after which the correct FES is found after removing the imposed symmetry.

Preliminary results of the metadynamics simulation show that the two coordination numbers, c_n(Cl−H), of the attacking ($c_n(\mathrm{Cl_A{-}H})$) and leaving ($c_n(\mathrm{Cl_L{-}H})$) chlorides visit distinct combinations during the progression along the main reaction coordinate. For example, in the stable reactant state, $c_n(\mathrm{Cl_A{-}H})$ and $c_n(\mathrm{Cl_L{-}H})$ fluctuate around (6,1), (5,1), (5,2), or (4,2). In the transition state, the coordination numbers fluctuate around (4,3), (3,3), (3,2), (2,2), (2,3), or (3,4). These distinct regions indicate that the rearrangement in the first coordination shells of the chlorides are indeed activated. Details on the population of these regions and the small barriers between them will be presented in a forthcoming paper. Here, we mainly want to illustrate the applicability of metadynamics to improve the description and sampling of reactions in complex systems.

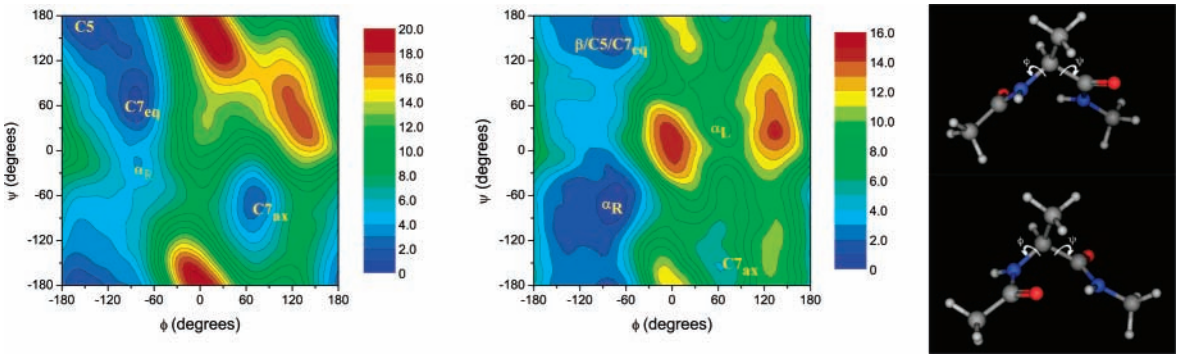

FIGURE 7. Free energy surfaces of the alanine dipeptide in gas phase (left) and in aqueous solution (middle) with a 1 kcal/mol contour interval. The right panel shows the dipeptide in the $C7_{eq}$ (top) and the $C7_{ax}$ (bottom) conformations.

Table 1. Relative Energies (kcal/mol) and ϕ/ψ Angles (deg, italic) of Minima on the Free Energy Surface of Alanine Dipeptide Obtained with Metadynamics Compared with Those from Selected Other Methods

	gas phase				aqueous solution			
	C5	$C7_{eq}$	α_R	$C7_{ax}$	β/C5/$C7_{eq}$	α_R	$C7_{ax}$	α_L
this work	0.24	0.0	3.18	1.15	1.0	0.0	4.8	7.4
	−153.2/175.2	*−84.3/64.7*	*−80.4/−17.8*	*70.6/−67.5*	*−85.3/167.1*	*−79.3/−71.4*	*62.1/−155.0*	*66.6/29.8*
FEP[a]	1.78	0.0		2.00	0.0	0.2	3.6	4.1
	−134.8/145.9	*−77.5/89.9*		*60.6/−72.4*	*−80/120*	*−80/−60*	*60/−80*	*60/60*
US[b]		0.0		2.6	0.3	0.0	5.0	7.5
		−74/70		*58/−49*	*−80/162*	*−72/−56*	*61/−133*	*59/57*
US[c]		0.0		2.7	0.0	1.41	3.85	4.38
		−74/99		*57/−83*	*−75/136*	*−76/−50*	*57/−84*	*51/81*
AFED[d]		0.0		2.3	0.2	0.0	4.5	
		−81/81		*63/−81*	*−81/153*	*−81/−63*	*63/−117*	
QM[e]	0.91	0.0	4.27	2.06				
	−158.5/151.5	*−81.9/81.6*	*−60/−45*[f]	*74.3/−57.3*				

[a] Free energy perturbation from ref 42. [b] Umbrella sampling from ref 43. [c] Umbrella sampling from ref 44. [d] Adiabatic free energy dynamics from ref 45. [e] Quantum mechanics calculation (MP2/6-311++G** geometries, LMP2/cc-VQZ(−g) energies) from ref 46. [f] ϕ/ψ dihedral angles constrained at −60/−45.

Note that if we would be interested in the FES of this reactive system completely treated QM (e.g., to study the difference between the MM versus QM treated solvent), we could save considerable time by first flattening the FES with metadynamics at the QM/MM level and then starting the metadynamics simulation of the fully QM system from the sum of hills already accumulated in the QM/MM simulation.

C. Metadynamics Compared to Other Methods: The Alanine Dipeptide Case. Our last application, the alanine dipeptide molecule (Figure 7) has long become a standard model system for theoretical studies of backbone conformational transitions of proteins. Various MD and Monte Carlo (MC) based free energy methods, including FEP, US, and the recently introduced adiabatic free energy dynamics (AFED) method, have been developed and applied on mapping the FES of alanine dipeptide as a function of the backbone dihedral angles ϕ and ψ.[5,7,12,42−45] In the present work, we have applied metadynamics to explore the 2D (ϕ,ψ) FES of alanine dipeptide in both gas and aqueous solution phases.

The alanine dipeptide and water molecules are described by the CHARMM27 force field[35] for proteins with the exception of modified intramolecular parameters for water. In the aqueous solution case, the alanine dipeptide molecule is placed into a periodic cubic box ($L = 18.8$ Å) with 216 water molecules. Computational details of this simulation and an error analysis are found in the Supporting Information. The total simulation times for the metadynamics runs were ∼5 ns for gas phase and ∼4 ns for aqueous solution phase, respectively.

A contour map of the FES of alanine dipeptide in gas phase is shown in Figure 7. The relative energies and positions (ϕ,ψ) of the minima are listed in Table 1, along with results of selected theoretical studies from published papers.

Regarding the comparison of gas phase FESs, the overall shapes of the free energy landscapes clearly evidence the two main minima $C7_{eq}$ and $C7_{ax}$, in good agreement with previous MD and QM data. However, Table 1 shows significant differences in the positions and relative energies of other minima, which are more shallow in nature and therefore hard to find, in particular: the C5 and α_R states. In fact, whereas three of the four other MD studies have missed the C5 minimum and none of them is able to find the α_R minimum, metadynamics successfully captures those minima. Our results show that the depth of the C5 minimum is more than 1 kcal/mol (larger than the estimated error of 0.3−0.5 kcal/mol) indicating that the C5 minimum is a reliable one. The α_R minimum is less certain (see for a more elaborate discussion the Supporting Information).

Also regarding the alanine dipeptide in aqueous solution, the overall features of the FES are in very good agreement among different methods. Moreover, they also agree well with the empirical Ramachandran plot con-

structed by Lovell et al., who used 81 234 residues from 500 high-resolution protein structures. All studies, including ours, indicate the stabilization of the folded α-helical conformations upon solvation. This feature results in a deeper α_R minimum with an energy as low as the extended β/C5/$C7_{eq}$ conformation. In the solvated case, two minima are found on the $\phi > 0$ side of the FES. One is the $C7_{ax}$ conformation; the other is the left-handed α_L conformation, which is stabilized by solvation and appears as a clear minimum. Thus, the overall order of the stability, starting from the most stable conformation, is isoenergetic α_R and β/C5/$C7_{eq}$ minima, then $C7_{ax}$ and, at last, α_L.

In conclusion, we believe to have shown that the metadynamics method is very good in accurately computing the FES. Compared to most of the other MD studies, metadynamics revealed more detail in less simulation time. Another strong point of metadynamics is that it does not require prior knowledge of the energy surface, contrary to other methods such as the FEP method or US techniques. For example, comparison of the adaptive US studies of Smith and Apostolakis shows differences of as much as 3 kcal/mol and 50° (see Table 1) due to differences in the creation of the biasing functions (Smith performed small numbers of iterations (three) with a long simulation (nanosecond scale) in each iteration, whereas Apostolakis used more iterations (200) with shorter simulations (50 ps)). Also, the FEP method utilizes simulation with constrained dihedral angles, thus requiring knowledge of a range of dihedral angles that encloses the minimum. In contrast, the parameters in the metadynamics method are relatively simple and can often be standardized in the study of multidimensional FESs.

IV. Perspective

Metadynamics has already revealed much of its potential in a wide range of applications, including ligand docking studies, crystal structure prediction, and chemical reactions.[25] The outlook is that development of extensions on the already mature metadynamics algorithm will further improve its efficiency, applicability and user friendliness. For example, parallelization via multiple walkers as well as tunable method parameters (such as the hills size and time interval) that adapt on the fly, learning from the collective variable dynamics along the FES exploration (see, for example, ref 29), reduce the simulation time and promise to bring the method closer to a tool that requires only minimal chemical intuition. Currently, we find that metadynamics in practice can effectively compute FESs as a function of up to three to four collective variables and accelerate the escape from local minima with up to four to six collective variables, depending on the system, which is enough to describe most chemical reactions. By exploiting symmetry in the FES, these numbers can be further increased. Metadynamics thus promises to soon become a standard method for the study of free energy landscapes of chemical reactions.

Supporting Information Available: The proof that the bias potential is an estimate of the free energy, metadynamics working recipe, computational details and error estimation of the alanine peptide simulations, discussion of whether the α_R minimum is a true minimum, and additional references regarding the alanine dipeptide and other uses of metadynamics. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic: San Diego, CA, 2002.

(2) Straatsma, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(3) Peters, B.; Heyden, A.; Bell A. T.; Chakraborty A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120*, 7877−7886.

(4) Zwanzig, R. W. High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(5) Mezei, M. Adaptive umbrella sampling − self-consistent determination of the nonboltzmann bias. *J. Comput. Phys.* **1987**, *68*, 237−248.

(6) Rebertus, D. W.; Berne, B. J.; Chandler, D. A molecular dynamics and Monte Carlo study of solvent effects on the conformational equilibrium of *n*-butane in CCl_4. *J. Chem. Phys.* **1979**, *70*, 3395−3400.

(7) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

(8) Kirkwood, G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.

(9) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.

(10) Gullingsrud, J. R.; Braun, R.; Schulten, K. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *J. Comput. Phys.* **1999**, *151*, 190−211.

(11) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.

(12) Olender, R.; Elber, R. Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* **1996**, *105*, 9299−9315.

(13) Passerone, D.; Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Phys. Rev. Lett.* **2001**, *87*, No. 108302.

(14) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, *108*, 1964−1977.

(15) Berg, B. A.; Neuhaus, T. Multicanonical algorithms for 1st order phase-transitions. *Phys. Lett. B* **1991**, *267*, 249−253.

(16) Sorensen, M. R.; Voter, A. F. Temperature-accelarated dynamics for simulations of infrequent events. *J. Chem. Phys.* **2000**, *112*, 9599−9606.

(17) Rosso, L.; Minary, P.; Zhu, Z.; Tuckerman, M. E. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.* **2002**, *116*, 4389−4402.

(18) He, J.; Zhang, Z.; Shi, Y.; Liu, H. Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *J. Chem. Phys.* **2003**, *119*, 4005−4017.

(19) Hooft, R. W. W.; Eijck, B. P. V.; Kroon, J. J. An adaptive umbrella sampling procedure in conformational-analysis using molecular dynamics and its application to glycol. *J. Chem. Phys.* **1992**, *97*, 6690−6694.

(20) Bartels, C.; Karplus, M. Multidimensional adaptive umbrella sampling: applications to main chain peptide conformations. *J. Comput. Chem.* **1997**, *18*, 1450−1462.

(21) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.

(22) Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **1995**, *52*, 2893−2906.

(23) Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.

(24) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(25) See the Supporting Information for a current listing of papers on metadynamics applications.

(26) Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. Microscopic mechanism of antibiotics translocation through a porin. *Biophys. J.* **2004**, *87*, 58−64.

(27) Wu, Y.; Schmitt, J. D.; Car, R. Mapping potential energy surfaces. *J. Chem. Phys.* **2004**, *121*, 1193−1200.
(28) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600−2607.
(29) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics. *Phys. Rev. Lett.* **2003**, *90*, No. 238302.
(30) Ensing, B., Klein, M. L. Perspective of the reactions between F^- and CH_3CH_2F: the free energy landscape of the E2 and S_N2 reaction channels. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6755−6759.
(31) Martoňák, R.; Laio, A.; Parrinello, M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **2003**, *90*, No. 075503.
(32) Donadio, D.; Raiteri, P.; Parrinello, M. Topological defects and bulk melting of hexagonal ice. *J. Phys. Chem. B* **2005**, *109*, 5421−5424.
(33) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the density of states by history-dependent metadynamics. *Phys. Rev. Lett.* **2004**, *92*, No. 170601.
(34) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B* **2005**, *109*, 6676.
(35) MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(36) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University: New York, 1989.
(37) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(38) Carloni, P.; Rothlisberger, U.; Parrinello, M. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* **2002**, *35*, 455−464.
(39) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714−6721.
(40) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical examination of the S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.
(41) Ensing, B.; Meijer, E. J.; Blöchl, P. E.; Baerends, E. J. Solvation effects on the S_N2 reaction between $CH_3Cl + Cl^-$, in water. *J. Phys. Chem. A* **2001**, *105*, 3300−3310.
(42) Tobias, D. J.; Brooks, C. L., III Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: a comparison of theoretical results. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(43) Smith, P. E. The alanine dipeptide free energy surface in solution. *J. Chem. Phys.* **1999**, *111*, 5568−5579.
(44) Apostolakis, J.; Ferrara, P.; Caflish, A. Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *J. Chem. Phys.* **1999**, *110*, 2099−2108.
(45) Rosso, L.; Abrams, J. B.; Tuckerman, M. E. Mapping the backbone dihedral free-energy surfaces in small peptides in solution using adiabatic free-energy dynamics. *J. Phys. Chem. B* **2005**, *109*, 4162−4167.
(46) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

AR040198I