Efficient and Direct Generation of Multidimensional Free Energy Surfaces via Adiabatic Dynamics without Coordinate Transformations

Jerry B. Abrams[†]

Department of Chemistry, New York University, New York, New York 10003

Mark E. Tuckerman*

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003

Received: June 8, 2008; Revised Manuscript Received: August 18, 2008

Adiabatic free energy dynamics (AFED) was introduced by Rosso et al. [*J. Chem. Phys.* **2002**, *116*, 4389] for computing free energy profiles quickly and accurately using a dynamical adiabatic separation between a set of collective variables or reaction coordinates and the remaining degrees of freedom of a system. This approach has been shown to lead to a significant gain in efficiency versus traditional methods such as umbrella sampling, thermodynamic integration, and free energy perturbation for generating one-dimensional free energy profiles. More importantly, AFED is able to generate multidimensional free energy surfaces efficiently via full sweeps of the surface that rapidly map out the locations of the free energy minima. The most significant drawback to the AFED approach is the need to transform the coordinates into a generalized coordinate system that explicitly contains the collective variables of interest. Recently, Maragliano and Vanden-Eijnden built upon the AFED approach by introducing a set of extended phase-space variables, to which the adiabatic decoupling and high temperature are applied [*Chem. Phys. Lett.* **2006**, *426*, 168]. In this scheme, which the authors termed "temperature accelerated molecular dynamics" or TAMD, the need for explicit coordinate transformations is circumvented. The ability of AFED and TAMD to generate free energy surfaces efficiently depends on the thermostatting mechanism employed, since both approaches are inherently nonequilibrium due to the adiabatic decoupling. Indeed, Maragliano and Vanden-Eijnden did not report any direct generation of free energy surfaces within the overdamped Langevin dynamics employed by these authors. Here, we show that by formulating TAMD in a manner that is closer to the original AFED approach, including the generalized Gaussian moment thermostat (GGMT) and multiple time-scale integration, multidimensional free energy surfaces for complex systems can be generated directly from the probability distribution function of the extended phase-space variables. The new TAMD formulation, which we term driven AFED or d-AFED, is applied to compare the conformational preferences of small peptides both in gas phase and in solution for three force fields. The results show that d-AFED/TAMD accurately and efficiently generates free energy surfaces in two collective variables useful for characterizing the conformations, namely, the radius of gyration, R_G, and number of hydrogen bonds, N_H.

1. Introduction

One of the key quantities in thermodynamics is the free energy associated with changes in conformation or the thermodynamic state of a complex system. Molecular dynamics (MD) and Monte Carlo based approaches have emerged as important theoretical tools to study such free energy changes along one-dimensional paths, e.g., along single reaction coordinates or one-dimensional λ-switching paths. Among these approaches, umbrella sampling[1−3] and thermodynamic integration[4−6] remain the most popular because they can be easily implemented. However, the problem of computing a multidimensional free energy surface (FES) in several collective variables or reaction coordinates of interest has remained a significant challenge, particularly when the FES contains numerous minima separated by high barriers. The mapping out of the free energy landscape of small peptides and proteins in the Ramachandran angles, radius of gyration and/or number of hydrogen bonds, or the characterization of dissociation or mass-transfer processes in aqueous solution in terms of coordination numbers and distances are examples of this type of problem.

The challenge of treating such "rough" energy landscapes has led to the introduction of various important new techniques for enhanced sampling of the configurational distribution of complex systems, from which the free energy is obtained. These include parallel tempering,[7−12] hyperdynamics,[13] parallel replica dynamics,[14] Wang−Landau sampling,[15,16] configuration-bias Monte Carlo,[17] the reference potential spatial-warping algorithm,[18,19] metadynamics,[20] and techniques based on adiabatic dynamics,[21−25] as a few examples.

The adiabatic free energy dynamics (AFED),[21−23] introduced eight years ago by Rosso et al.,[26] is a dynamical scheme for generating free energy hypersurfaces in several collective variables of interest. The approach employs an imposed adiabatic decoupling between a small set of collective variables or reaction coordinates and the remaining degrees of freedom. Within this scheme, an elevated temperature is also applied to the collective variables to ensure that they are able to cross the high-energy barriers needed to ensure sufficient sampling. In the limit of high temperature and adiabatic decoupling, it can be shown that the free energy hypersurface in the collective variables is

* To whom correspondence should be addressed.
[†] Current address: Department of Chemistry, Boston University, Boston, MA 02215.

10.1021/jp805039u CCC: $40.75
Published on Web 11/12/2008

obtained directly from their resultant probability distribution function.[21,22] The approach has been applied to the conformational sampling of small peptides[23] and in the computation of solvation and binding free energies via alchemical transformations.[25] In both cases, the use of adiabatic dynamics has been shown to lead to a significant improvement in efficiency compared to traditional methods such as free energy perturbation,[27] umbrella sampling,[1−3] and the blue moon ensemble approach.[5,6] In addition to being a relatively fast method, the AFED approach requires no a posteriori processing of the simulation data. Moreover, AFED is able to generate multidimensional free energy hypersurfaces with significantly greater efficiency than multidimensional versions of the aforementioned approaches.[23] By construction, AFED generates full sweeps of the free energy surface and, therefore, can rapidly map out the locations of the free energy minima well before the entire surface is fully converged.

The AFED approach is derived and implemented, in practice, by transforming the coordinate integrations in the canonical partition function to a set of generalized coordinates that explicitly contain the collective variables of interest. This gives rise to the disadvantage that the adiabatic dynamics must be carried out in these generalized coordinates, which leads to a steep implementation curve due to the rather invasive modifications to existing MD packages needed to introduce these transformations. It should be noted, however, that once such transformations are put in place, they can be subsequently combined with additional spatial-warping transformations that also significantly enhance conformational sampling.[18,19]

Recently, Maragliano and Vanden-Eijnden built on the AFED approach by introducing a set of extended phase-space or "driving" variables that are harmonically coupled to the collective variables of interest.[24] By imposing the adiabatic decoupling and high temperature on these extended variables rather than on the collective variables, the need for explicit transformations is avoided, thereby enlarging the class of collective variables that can be treated and rendering the technique substantially easier to implement. Maragliano and Vanden-Eijnden named the new technique "temperature accelerated molecular dynamics" or TAMD. It should be noted that such "driving" variables are also central in the so-called "metadynamics" approach,[20] where they are used together with a time-dependent potential that floods energy basins with Gaussians, thereby allowing the system to escape the basin and move into a neighboring one. In metadynamics, as the basins are filled, the histogram in the collective variables becomes flat. When this occurs, the sum of all of the Gaussians is used to recover the free energy hypersurface.

In contrast to metadynamics, the principal advantage of the adiabatic dynamics/TAMD approach lies in the fact that the elevated temperature of the collective variables allows the general shape of a free energy surface, in particular, the location of the minima, to be mapped out rapidly and with greater efficiency than is required to fully converge the surface. By contrast, in the metadynamics approach, each basin must be filled completely before the system can move to an adjacent basin, although the use of multiple walkers has been shown to alleviate this problem.[28] It is worth pointing out that multiple simulations initiated in different energy basins could be implemented straightforwardly in parallel in the AFED approach as a means of further increasing its sampling efficiency. In general, filling each location on the free energy surface with a Gaussian, while efficient, leads to a situation where it is possible, and even common, that the lowest energy regions might only be studied last and after a great deal of computational effort. Additionally, as the dimensionality of the system increases, the amount of time required for a metadynamics calculation to move into distant regions of the surface increases exponentially. The adiabatic/TAMD schemes, however, scan the entire free energy surface in full sweeps due to the elevated temperature, and consequently, minima are located rapidly, despite the increase with dimensionality of the time needed to converge the surface in its entirety. Finally, the AFED and TAMD schemes have fewer parameters than metadynamics.

Maragliano and Vanden-Eijnden formulated the TAMD approach in terms of overdamped Langevin dynamics and did not report any actual free energy surfaces.[24] Rather, these authors recently employed TAMD to provide an initial "sweep" of the free energy surface in order to identify important regions.[29] From this initial exploration of the surface, a variational technique is then used to reconstruct the surface in greater detail. This novel approach can also be used in conjunction with other techniques capable of generating an initial sweep of a free energy surface.

The purpose of the present paper is to provide a reformulation of TAMD that is closer in spirit to the original AFED approach, wherein it will be shown that probability distributions and free energy hypersurfaces in at least two dimensions for systems of moderate complexity can be generated *directly* from an adiabatic dynamics simulation. The reformulation we propose, which we term "driven AFED" or d-AFED, employs nonequilibrium extended-system thermostating techniques[30] and multiple-time-scale integration[32] on the harmonic coupling term. Through applications that include the alanine dipeptide, *N*-acetyl-tryptophan-methylamide (NATMA), *N*-acetyl-tryptophan-amide (NATA), and *N*-acetyl-$(Ala)_6$-methylamide, we show that our reformulation of TAMD is capable of rapidly locating the free energy minima and efficiently generating fully converged free energy hypersurfaces in a two-dimensional collective variable space consisting of the radius of gyration and number of hydrogen bonds.

This paper is organized as follows: A brief review of the AFED approach is presented in section 2. In section 3, a derivation of the new d-AFED/TAMD approach is given, and a model problem is used to illustrate how the new d-AFED/TAMD approach works. In section 4, the results of the d-AFED simulations of the alanine dipeptide in gas phase, and in solution, are presented. Conformational preferences are compared with previous theoretical studies, and the efficiency of the method is compared to traditional MD. Additionally, the advantages of using r-RESPA factorization in d-AFED is examined and presented. In sections 5 and 6, a comparison of the results for the conformational sampling of *N*-acetyl-tryptophan-methylamide (NATMA) and *N*-acetyl-tryptophan-amide (NATA), in gas phase for the three major force fields, are presented, respectively. Preliminary results for the conformational sampling of *N*-acetyl-$(Ala)_6$-methylamide (alanine hexamer) in solution are presented in section 7. Conclusions are given in section 8.

2. Adiabatic Free Energy Dynamics

Consider a system of N particles with Cartesian coordinates $\mathbf{r}_1,...,\mathbf{r}_N \equiv \mathbf{r}$ and conjugate momenta $\mathbf{p}_1,...,\mathbf{p}_N \equiv \mathbf{p}$ subject to a potential energy $V(\mathbf{r}_1,...,\mathbf{r}_N)$. The classical canonical partition function for the system is given by

$$Q = C \int \mathrm{d}^N\mathbf{p} \int_{D(V)} \mathrm{d}^N\mathbf{r} \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(\mathbf{r}_1, ..., \mathbf{r}_N)\right]\right\} \quad (1)$$

where $H(\mathbf{p},\mathbf{r}) = \sum_i \mathbf{p}_i^2/2m_i + V(\mathbf{r})$ is the Hamiltonian, $\beta = 1/k_B T$, $D(V)$ is the spatial domain defined by the containing volume, and C is an overall prefactor that renders Q dimensionless and compensates for overcounting of states obtained by exchanging particles of the same chemical identity. For a system with M species, $C = [h^{3N}\Pi_{\alpha=1}^{M} N_\alpha!]^{-1}$, where h is Planck's constant and N_α is the number of particles of species α.

Suppose we wish to determine the free energy hypersurface in a set of $n < N$ collective variables $q_1(\mathbf{r}),..., q_n(\mathbf{r})$. Examples are the Ramachandran angles for characterizing the conformational space of oligopeptides or combinations of distances for characterizing a chemical reaction. The probability that $q_1(\mathbf{r})$ has the value s_1, $q_2(\mathbf{r})$ has the value s_2,..., $q_n(\mathbf{r})$ has the value s_n is given by

$$P(s_1, ..., s_n) = \frac{\int d^N\mathbf{p}\, d^N\mathbf{r}\, e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{i=1}^{n} \delta(q_i(\mathbf{r}) - s_i)}{\int d^N\mathbf{p}\, d^N\mathbf{r}\, e^{-\beta H(\mathbf{p},\mathbf{r})}} \quad (2)$$

Given this probability distribution, the free energy hypersurface can be calculated according to

$$F(s_1, ..., s_n) = -kT \ln P(s_1, ..., s_n) \quad (3)$$

In many complex systems, direct calculation of the probability distribution function from a molecular dynamics trajectory is intractable because of the existence of high potential energy barriers separating important minima on the hypersurface. Potential energy surfaces of this type are said to be "rough", and it is necessary to employ enhanced sampling techniques. The adiabatic free energy dynamics (AFED) achieves enhanced sampling in the variables $q_1(\mathbf{r}),..., q_n(\mathbf{r})$ by introducing a high temperature $T_s \gg T$ for these n degrees of freedom only, while maintaining the remaining $3N - n$ degrees of freedom at the correct ensemble temperature T. The temperature disparity can be accomplished by introducing two separate sets of thermostats for each set of degrees of freedom. The high temperature T_s ensures that the variables $q_1,..., q_n$ are able to cross high-energy barriers on their part of the energy landscape. However, this high temperature also destroys the thermodynamic properties of the system *unless* the variables $q_1,..., q_n$ are also adiabatically decoupled from the remaining degrees of freedom. In order to accomplish this decoupling, we need to be able to run the dynamics in a coordinate system that explicitly contains $q_1,..., q_n$.

Suppose there is a transformation from Cartesian coordinates $\mathbf{r}_1,.., \mathbf{r}_N$ to generalized coordinates $q_1,..., q_{3N} \equiv q$ via the transformation equations $q_\alpha = q_\alpha(\mathbf{r})$. The inverse transformations are denoted $\mathbf{r}_i = \mathbf{r}_i(q)$. Substituting the transformation into eq 1 yields

$$Q = C \int d^N\mathbf{p} \int_{D(V)} d^{3N}q J(q) \times \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}_1(q), ..., \mathbf{r}_N(q))\right]\right\} = C \int d^N\mathbf{p} \int_{D(V)} d^{3N}q \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \tilde{V}(q_1, ..., q_{3N})\right]\right\} \quad (4)$$

where the potential $\tilde{V}$ contains the Jacobian of the transformation $J(q) = |\partial\mathbf{r}/\partial q|$ and is given by

$$\tilde{V}(q_1, ..., q_{3N}) = V(\mathbf{r}_1(q), ..., \mathbf{r}_N(q)) - kT \ln J(q_1, ..., q_{3N}) \quad (5)$$

Note that the partition function in eq 4 is completely equivalent to that in eq 1. Moreover, even though the transformation is not canonical, since we are only interested in sampling the ensemble distribution, we can treat the $3N$ Cartesian momentum components as "conjugate" to the $3N$ generalized coordinates, each being *defined* as $p_\alpha = m_\alpha \dot{q}_\alpha$, $\alpha = 1,.., 3N$, where m_α are the associated masses. Thus, in order to achieve the desired adiabatic decoupling, we simply choose the first n masses m_α to be much larger than all of the remaining masses, $m_{1,..., n} \gg m_{n+1,...,3N}$.

Under the conditions of adiabatic decoupling and the temperature disparity, it was shown in refs 21 and 22 via a decomposition of the classical propagator that the probability distribution, denoted $P_{adb}(s_1,...,s_n)$, becomes

$$P_{adb}(s_1, ..., s_n) = N \int d^n p \exp\left[-\beta_s \sum_{\alpha=1}^{n} \frac{p_\alpha^2}{2m_\alpha}\right] [Z(s_1, ..., s_n, \beta)]^{T/T_s} \quad (6)$$

where

$$Z(s_1, ..., s_n, \beta) = \int d^{3N-n}p\, d^{3N}q \times \exp\left\{-\beta\left[\sum_{\alpha=n+1}^{3N} \frac{p_\alpha^2}{2m_\alpha} + \tilde{V}(q_1, ..., q_{3N})\right]\right\} \prod_{\alpha=1}^{n} \delta(q_\alpha - s_\alpha) \quad (7)$$

and $\mathbf{N}$ is an overall normalization factor. In this case, because of the temperature ratio T/T_s in the exponent, the exact free energy $F(s_1,...,s_n)$ at the temperature T, which is defined to be $F(s_1,...,s_n) = -kT \ln Z(s_1,...,s_n,\beta)$, is obtained from $P_{adb}(s_1,...,s_n)$ by

$$F(s_1, ..., s_n) = -kT_s \ln P_{adb}(s_1, ..., s_n) \quad (8)$$

Note that the multiplicative factor $-kT_s$ in eq 8 ensures that the free energy *at temperature T* is obtained. Equation 8 shows that the free energy surface can be computed *directly* from the probability distribution function generated in an adiabatic dynamics calculation.

We now briefly review the derivation of eqs 7 and 8. Readers already familiar with the derivation or wishing to skip ahead can jump to the end of this section without loss of continuity. In order to simplify the notation, let $\tilde{q} = (q_1,.., q_n)$ denote the first n collective variables of interest and $q = (q_{n+1},..., q_{3N})$ denote the remaining generalized coordinates. Let $\tilde{p}$ and p denote the corresponding momentum vectors and $\tilde{M}$ and M denote the corresponding diagonal mass matrices. The adiabatic Hamiltonian then takes the form

$$\tilde{H} = \frac{1}{2}\tilde{p}^T\tilde{M}^{-1}\tilde{p} + \frac{1}{2}p^T M^{-1} p + \tilde{V}(\tilde{q}, q) \quad (9)$$

The equations of motion derived from eq 9 must be coupled to two sets of thermostats in order to maintain the two temperatures

T_s and T. Assuming a dynamical thermostating scheme in an extended phase space such as the Nosé−Hoover chain[74] or generalized Gaussian moment[30] approach, the Liouville operator for the dynamical propagation of the system can be written in the form

$$iL = \tilde{M}^{-1}\tilde{p}\cdot\frac{\partial}{\partial\tilde{q}} + M^{-1}p\cdot\frac{\partial}{\partial q} + F_{\tilde{q}}\cdot\frac{\partial}{\partial\tilde{p}} + F_q\cdot\frac{\partial}{\partial p} + iL^{(\mathrm{th})}(\tilde{p},T_s) + iL^{(\mathrm{th})}(p,T) \quad (10)$$

where $F_{\tilde{q}}$ and F_q are the forces on the $\tilde{q}$ and q coordinates, respectively, and $iL^{(\mathrm{th})}(\tilde{p},T_s)$ and $iL^{(\mathrm{th})}(p,T)$ are the Liouville operators for the thermostats that maintain the temperatures T_s and T, respectively. The thermostats are assumed to affect the physical momenta only, which explains the dependence of these operators on $\tilde{p}$ and p. Using the Liouville operator, we seek a factorization of the classical propagator $\exp(iLt)$ appropriate for analyzing the dynamics. To this end, we define the following contributions to the Liouville operator

$$iL_{\tilde{q}}^{(\mathrm{ref})} = \tilde{M}^{-1}\tilde{p}\frac{\partial}{\partial\tilde{q}} + iL^{(\mathrm{th})}(\tilde{p},T_s)$$

$$iL_q^{(\mathrm{ref})} = M^{-1}p\frac{\partial}{\partial q} + iL^{(\mathrm{th})}(p,T)$$

$$iL_q = iL_q^{(\mathrm{ref})} + F_{\tilde{q}}\frac{\partial}{\partial\tilde{p}} + F_q\frac{\partial}{\partial p} \quad (11)$$

so that $iL = iL_q + iL_{\tilde{q}}^{(\mathrm{ref})} + iL_q^{(\mathrm{ref})}$. We now choose a time interval Δt appropriate for discretizing the motion of the slow variables $\tilde{q}$ and factorize the single time-step propgator $\exp(iL\Delta t)$ using the Trotter theorem according to $\exp(iL\Delta t) = \exp(iL_q\Delta t/2)\exp(iL_{\tilde{q}}^{(\mathrm{ref})}\Delta t)\exp(iL_q\Delta t/2) + O(\Delta t^3)$. Since the time interval Δt is long compared to the motion of the fast variables q, the operator $\exp(iL_q\Delta t/2)$ needs to be further factorized. We choose an exact representation of this operator based on the Trotter theorem:

$$\exp\left(iL_q\frac{\Delta t}{2}\right) = \lim_{n\to\infty}\left[\exp\left(\frac{\Delta t}{4n}F_{\tilde{q}}\cdot\frac{\partial}{\partial\tilde{p}}\right)\exp\left(\frac{\Delta t}{4n}F_q\cdot\frac{\partial}{\partial p}\right)\times \exp\left(iL_q^{(\mathrm{ref})}\frac{\Delta t}{2n}\right)\exp\left(\frac{\Delta t}{4n}F_q\cdot\frac{\partial}{\partial p}\right)\exp\left(\frac{\Delta t}{4n}F_{\tilde{q}}\cdot\frac{\partial}{\partial\tilde{p}}\right)\right]^n \quad (12)$$

When the factorized propagator $\exp(iL\Delta t)$ is used to propagate the phase space variables, the force acting on the slow variables is a time average of $F_{\tilde{q}}$ over the motion of the fast variables for an interval $\Delta t/2$[31,21,22]

$$\bar{F}_{\tilde{q}} = \frac{2}{\Delta t}\int_\tau^{\tau+\Delta t/2} \mathrm{d}t\, F_{\tilde{q}}(q(t;x(\tau)),\tilde{q}) \quad (13)$$

where $x(\tau)$ is the full phase space vector at time τ and $\tau = 0$ or $\Delta t/2$. The integral arises as follows: The repeated action of eq 13 on the phase space generates a sum n force terms in the evolution of the slow variables in which $F_{\tilde{q}}$ is evaluated at fixed $\tilde{q}$ but time-varying q, with q evaluated a times $\Delta t/2n$ in its evolution. The sum is the trapezoidal rule representation of the time integral, which becomes a continuous integral when $n \to \infty$. Finally, we assume that q visits enough of its available phase space over the interval Δt of the slow variables that the time average can be replaced by a phase space average:

$$\frac{2}{\Delta t}\int_\tau^{\tau+\Delta t/2} \mathrm{d}t\, F_{\tilde{q}}(q(t;x(\tau)),\tilde{q}) = \frac{\partial}{\partial\tilde{q}}\frac{1}{\beta}\ln Z_q(\tilde{q},\beta) \quad (14)$$

where

$$Z_q(\tilde{q},\beta) = \int \mathrm{d}q\, \mathrm{e}^{-\beta\tilde{V}(\tilde{q},q)} \quad (15)$$

Thus, we see that, under the assumption of ergodicity, the adiabatic dynamics generates the potential of mean force $-kT\ln Z_q(\tilde{q},\beta)$ for the slow variables "on the fly". We can now write down an effective Hamiltonian for the slow variables $H_{\mathrm{eff}}(\tilde{p},\tilde{q}) = \tilde{p}^{\mathrm{T}}\tilde{M}^{-1}\tilde{p}/2 - kT\ln Z_q(\tilde{q},\beta)$. Since the slow variables are independently thermostated, their dynamics will sample a canonical distribution in H_{eff} at temperature T_s

$$\tilde{Q}(T_s,T) = \tilde{C}\int \mathrm{d}\tilde{p}\,\mathrm{d}\tilde{q}\,\mathrm{e}^{-\beta_s H_{\mathrm{eff}}(\tilde{p},\tilde{q})} = \tilde{C}\int \mathrm{d}\tilde{p}\,\mathrm{e}^{-\beta_s\tilde{p}^{\mathrm{T}}\tilde{M}^{-1}\tilde{p}/2}\int \mathrm{d}\tilde{q}\,[Z_q(\tilde{q},\beta)]^{\beta_s/\beta} \quad (16)$$

which is equivalent to eq 7 if the identification $\tilde{q}_1 = s_1,\ldots,\tilde{q}_n = s_n$ is made. The probability that $\tilde{q}_1 = s_1,\ldots,\tilde{q}_n = s_n$ is, therefore, given by eq 6, and the free energy surface, defined to be $F(s_1,\ldots,s_n) = -kT\ln Z_q(s_1,\ldots,s_n,T)$ is, therefore, given by eq 8.

The AFED procedure can be summed up as follows:

1. Choose the temperature T_s and the first n masses $m_1,\ldots,m_n$. Generally, T_s should be chosen high enough to ensure that all barriers in the variables $q_1,\ldots,q_n$ can be crossed. The masses $m_1,\ldots,m_n$ can be chosen as some overall multiplicative factor g times a typical mass in the system, e.g., the mass of a carbon or oxygen atom. This gives just two parameters to set. g needs to be chosen high enough to ensure the adiabatic decoupling at the chosen T_s.

2. Use the Hamiltonian

$$\tilde{H}(p,q) = \sum_{\alpha=1}^{3N}\frac{p_\alpha^{\,2}}{2m_\alpha} + \tilde{V}(q_1,\ldots,q_{3N}) \quad (17)$$

to derive the equations of motion.

3. Couple the equations of motion to two separate thermostats, the first n being at temperature T_s and the remaining at temperature T.

4. Run the adiabatic dynamics to accumulate the probability distribution $P_{\mathrm{adb}}(s_1,\ldots,s_n)$ from a simple histogram, and use eq 8 to obtain the free energy surface.

3. Adiabatic Free Energy Dynamics without Transformations

The AFED approach is a powerful one that is capable of generating multidimensional free energy surfaces efficiently, as was shown in refs 22 and 23. However, the need to work in generalized coordinates is a distinct disadvantage of the method, as this requires rather invasive modifications to existing molecular dynamics codes.

Recently, Maragliano and Vanden-Eijnden[24] showed that AFED could be re-expressed in a set of extended phase space

variables in a manner similar to that used in the metadynamics approach of Laio and Parrinello,[20] thereby circumventing the need for explicit coordinate transformations. This new formulation, which the authors called "temperature accelerated molecular dynamics" (TAMD), increases both the flexibility of the AFED method, allowing larger classes of collective variables to be treated, and the ease of implementation in existing packages.

TAMD can be derived as follows. We rewrite the product of δ-functions in eq 2 as the limit of a product of Gaussian functions[33]

$$\prod_{\alpha=1}^{n}\delta(q_\alpha(\mathbf{r})-s_\alpha)=\lim_{\kappa\to\infty}\sqrt{\frac{\beta k}{2\pi}}\times \exp\left[-\sum_{\alpha=1}^{n}\frac{\beta}{2}k(q_\alpha(\mathbf{r})-s_\alpha)^2\right] \quad (18)$$

When eq 18 is substituted into eq 2, we obtain

$$P(s_1,...,s_n)=\lim_{\kappa\to\infty}N_\kappa\int \mathrm{d}^N\mathbf{p}\int_{D(V)}\mathrm{d}^N\mathbf{r}\exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r})+\frac{1}{2}\kappa\sum_{\alpha=1}^{n}(q_\alpha(\mathbf{r})-s_\alpha)^2\right]\right\} \quad (19)$$

where $\mathbf{N}_\kappa$ is a κ-dependent normalization constant. For large but finite κ, the integral in eq 19 represents a close approximation to the true probability distribution, and we can regard the harmonic term in eq 19 as an additional potential term that keeps the collective variables $q_1(\mathbf{r}),...,q_n(\mathbf{r})$ close to the values $s_1,...,s_n$. In this representation, eq 19 resembles the probability distribution generated within the umbrella sampling approach.[1−3] However, if a set of n independent Gaussian integrations is introduced into eq 19 in the following form

$$P(s_1,...,s_n)=\lim_{\kappa\to\infty}N'_\kappa\int \mathrm{d}^n p_s\int \mathrm{d}^N\mathbf{p}\int_{D(V)}\mathrm{d}^N\mathbf{r}\times \exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r})+\sum_{\alpha=1}^{n}\frac{{p_{s_\alpha}}^2}{2m_\alpha}+\frac{1}{2}\kappa\sum_{\alpha=1}^{n}(q_\alpha(\mathbf{r})-s_\alpha)^2\right]\right\} \quad (20)$$

then the dependence of the distribution on $s_1,...,s_n$ remains unaltered.

The argument of the exponential can now be regarded as an extended phase space Hamiltonian

$$H_{\mathrm{ex}}(\mathbf{p},p_s,\mathbf{r},s)=\sum_{\alpha=1}^{n}\frac{{p_{s_\alpha}}^2}{2m_\alpha}+\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}+V(\mathbf{r}_1,...,\mathbf{r}_N)+\sum_{\alpha=1}^{n}\frac{1}{2}\kappa(q_\alpha(\mathbf{r})-s_\alpha)^2 \quad (21)$$

This Hamiltonian generates the dynamics of the original N Cartesian positions and momenta and of the additional n variables $s_1,...,s_n \equiv s$ and their conjugate momenta $p_{s_1},...,p_{s_n} \equiv p_s$. The extended variables serve to "drag" or "drive" the collective variables $q_1(\mathbf{r}),...,q_n(\mathbf{r})$ via the harmonic coupling through their portion of the energy landscape provided that the variables $s_1,...,s_n$ are able to sample a comparable region.

Assuming, again, that there are significant barriers hindering the sampling of the collective variables, enhanced sampling can be achieved by employing a high temperature and adiabatic decoupling, this time on the extended phase space variables.[24] Thus, we introduce a temperature $T_s \gg T$ and masses $m_\alpha \gg m_i$ for these variables. As in the original AFED scheme, the former condition ensures that high barriers can be crossed, if T_s is chosen high enough, while the large masses ensure adiabatic decoupling of the extended phase space variables from all other degrees of freedom. Following refs 21, 22, and 25, it can be shown that, under these conditions, the distribution function generated takes the form

$$P^{(\kappa)}_{\mathrm{adb}}(s_1,...,s_n)\propto\int \mathrm{d}^n p_s\exp\left[-\beta_s\sum_{\alpha=1}^{n}\frac{{p_{s_\alpha}}^2}{2m_\alpha}\right][Z(s_1,...,s_n,\beta)]^{\beta_s/\beta} \quad (22)$$

where $\beta_s = 1/kT_s$ and

$$Z(s_1,...,s_n,\beta)=\int \mathrm{d}^N\mathbf{p}\int_{D(V)}\mathrm{d}^N\mathbf{r}\exp\left\{-\beta\left[\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}+\bar{V}(\mathbf{r},s)\right]\right\} \quad (23)$$

and

$$\bar{V}(\mathbf{r},s)=V(\mathbf{r}_1,...,\mathbf{r}_N)+\frac{1}{2}\kappa\sum_{\alpha=1}^{n}(q_\alpha(\mathbf{r})-s_\alpha)^2 \quad (24)$$

The probability distribution in eq 22 generates an approximation $F_\kappa(s_1,...,s_n)$ to the true free energy profile at temperature T according to

$$F_\kappa(s_1,...,s_n)=-kT_s\ln P^{(\kappa)}_{\mathrm{adb}}(s_1,...,s_n) \quad (25)$$

and it is clear that in the limit $\kappa\to\infty$ the true free energy profile is recovered

$$F(s_1,...,s_n)=\lim_{\kappa\to\infty}F_\kappa(s_1,...,s_n) \quad (26)$$

Equations 25 and 26 show that the free energy hypersurface can be generated within the adiabatic dynamics scheme without requiring a transformation to generalized coordinates.

The ability of TAMD to generate the free energy surface efficiently depends on the thermostatting mechanism employed to maintain the two temperatures. The adiabatic decoupling represents a nonequilibrium steady state, and in refs 21 and 22 it was shown that the generalized Gaussian moment thermostat (GGMT) of Liu and Tuckerman[30] is an effective approach for maintaining the temperature disparity within the AFED scheme. Therefore, we employ it here as well. For completeness, we show the explicit equations of motion, including the coupling to separate GGMTs at temperatures T and T_s. As noted, GGMTs are capable of maintaining temperature control under the nonequilibrium (steady-state) conditions implied by the two

temperatures and adiabatic decoupling. Within the two-moment version of the GGMT technique, with a separate thermostat coupled to each degree of freedom, the equations of motion for the new TAMD scheme read

$$\dot{r}_{i,k} = \frac{p_{i,k}}{m_i}$$

$$\dot{p}_{i,k} = F_{i,k} - \kappa \sum_{\alpha=1}^{n} (q_\alpha(\mathbf{r}) - s_\alpha)\frac{\partial q_\alpha}{\partial r_{i,k}} - \frac{p_{\eta_{i,k,1}}}{Q_1}p_{i,k} - \frac{p_{\eta_{i,k,2}}}{Q_2}\left[(kT)p_{i,k} + \frac{{p_{i,k}}^3}{3m_i}\right]$$

$$\dot{s}_\alpha = \frac{p_{s_\alpha}}{m_\alpha}$$

$$\dot{p}_{s_\alpha} = \kappa(q_\alpha(\mathbf{r}) - s_\alpha) - \frac{p_{\xi_{\alpha,1}}}{Q'_1}p_{s_\alpha} - \frac{p_{\xi_{\alpha,2}}}{Q'_2}\left[(kT_s)p_{s_\alpha} + \frac{{p_{s_\alpha}}^3}{3m_\alpha}\right]$$

$$\dot{\eta}_{i,k,1} = \frac{p_{\eta_{i,k,1}}}{Q_1}$$

$$\dot{\eta}_{i,k,2} = \left[(kT) + \frac{{p_{i,k,1}}^2}{m_i}\right]\frac{p_{\eta_{i,k,2}}}{Q_2}$$

$$\dot{\xi}_{\alpha,1} = \frac{p_{\xi_{\alpha,1}}}{Q'_1}$$

$$\dot{\xi}_{\alpha,2} = \left[(kT_s) + \frac{{p_{s_\alpha}}^2}{m_\alpha}\right]\frac{p_{\xi_{\alpha,2}}}{Q'_2}$$

$$\dot{p}_{\eta_{i,k,1}} = \frac{{p_{i,k}}^2}{m_i} - kT$$

$$\dot{p}_{\eta_{i,k,2}} = \frac{{p_{i,k}}^4}{3{m_i}^2} - (kT)^2$$

$$\dot{p}_{\xi_{\alpha,1}} = \frac{{p_{s_\alpha}}^2}{m_\alpha} - kT_s$$

$$\dot{p}_{\xi_{\alpha,2}} = \frac{{p_{s_\alpha}}^4}{3{m_\alpha}^2} - (kT_s)^2 \qquad (27)$$

where $F_{i,k} = -\partial V/\partial r_{i,k}$. In eqs 27, the thermostats are used to control the fluctuations in the second and fourth moments of the distribution of each momentum variable in the system, whether these correspond to T or T_s. Here, k indexes the three Cartesian components of the physical coordinate and momentum of particle i, and $\eta_{i,k,1}$, $\eta_{i,k,2}$, $p_{\eta_{i,k,1}}$, and $p_{\eta_{i,k,2}}$ are the corresponding GGMT variables. Similarly, α indexes the extended phase space driving variables, and $\xi_{\alpha,1}$, $\xi_{\alpha,2}$, $p_{\xi_{\alpha,1}}$, and $p_{\xi_{\alpha,2}}$ are the corresponding GGMT variables. The thermostat mass paramters Q_1, Q_2, Q'_1, and Q'_2 are chosen according to[30]

$$Q_1 = kT\tau^2 \qquad Q_2 = \frac{8}{3}(kT)^3\tau^2$$

$$Q'_1 = kT_s{\tau_s}^2 \qquad Q'_2 = \frac{8}{3}(kT_s)^3{\tau_s}^2 \qquad (28)$$

where τ and τ_s are characteristic time scales in the physical and extended systems, respectively. A typical choice for τ_s is the period of the harmonic coupling $(2\pi)(m_\alpha/\kappa)^{1/2}$.

Another important feature of eqs 27 is that the presence of the stiff harmonic force term $(\kappa/2)\sum_\alpha(q_\alpha(\mathbf{r}) - s_\alpha)^2$ renders them amenable to multiple time scale (r-RESPA) integration techniques.[32,34] For eqs 27, the Liouville operator, from which such integrators are derived, can be subdivided into a reference system containing the stiff oscillations of the harmonic coupling and a second propagator for the relatively slow motions associated with the motion of the physical system. Using r-RESPA to evaluate the inexpensive, but very fast, stiff harmonic force with a smaller time step, while keeping the fundamental time step larger, significantly improves the efficiency of the method. For details of this type of factorization the interested reader is directed to refs 32 and 34. Although Maragliano and Vanden-Eijnden named the extended phase space approach "temperature accelerated molecular dynamics" (TAMD),[24] we prefer to retain the AFED acronym, as it is more descriptive of the nature of the adiabatic dynamics employed here. Thus, we name the scheme in eqs 27 combined with multiple time-scale integration the driven AFED or d-AFED approach, although we will continue to refer to both names. Maragliano and Vanden-Eijnden discuss the use of TAMD as an approach for the initial exploration of the free energy surface,[29] following which the authors use a variational technique for reconstructing the surface in greater detail. However, no actual free energy surfaces using TAMD were reported. In the subsequent sections, we show that eqs 27 with r-RESPA integration are capable of generating surfaces in multiple dimensions for small oligopeptides, in gas phase and in solution, directly from the adiabatic dynamics.

Before moving on to these application, however, we illustrate how the d-AFED/TAMD method works on a simple example problem that was first introduced in ref 22. The system is a one-dimensional quartic double well coupled to a harmonic oscillator described by two degrees of freedom, x and y, subject to a potential of the form

$$V(x, y) = D_0(x^2 - a^2)^2 + \frac{1}{2}\kappa y^2 + \lambda xy \qquad (29)$$

For this example problem, the configurational partition function, $Z_y(x;\beta_y)$, can be calculated analytically, and the resulting symmetric free energy surface in x is given by

$$F(x) = -\frac{1}{\beta_y}\ln Z_y(x;\beta_y) = D_0(x^2 - a^2)^2 - \frac{\lambda^2}{2\kappa}x^2 \qquad (30)$$

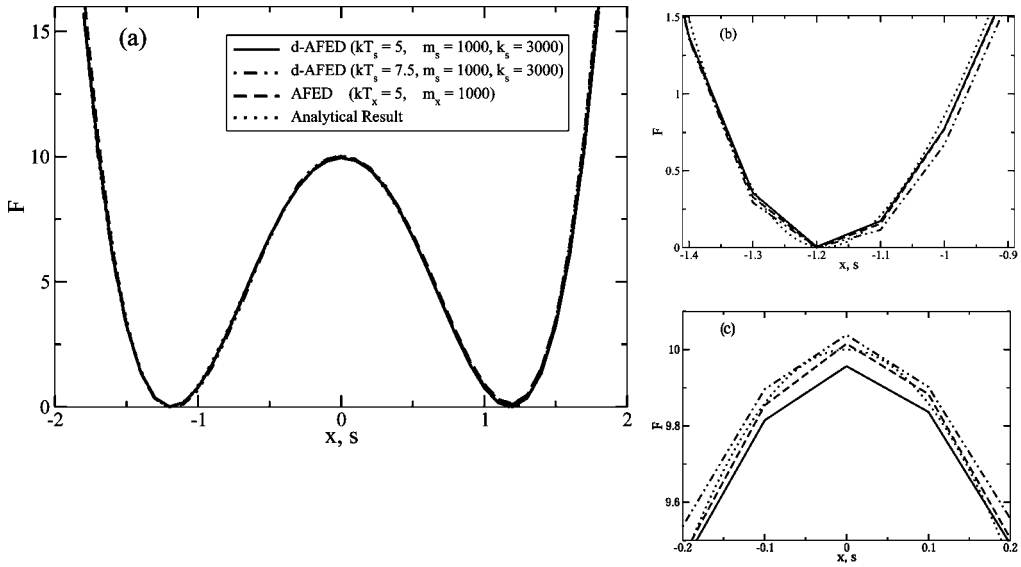

Figure 1. Free energy profiles computed for the one-dimensional quartic double well coupled to a harmonic oscillator using AFED (dashed lines) and d-AFED/TAMD (solid and dot-dashed lines). The analytical result is presented with dotted lines. (a) Comparison of the entire free energy profiles, (b) zoomed-in look at one of the minima, and (c) zoomed-in look at the barrier region.

AFED simulations of this model were performed using the same parameters that were used in ref 22: $D_0 = 5$, $a = 1$, $\kappa = 1$, $\lambda = 2.878$, $m_y = 1$, $kT_y = 1$, and $\Delta t = 0.25 \times 10^{-3}$. The resultant free energy profiles in x minima at $x_\pm = \pm(a^2 + \lambda^2/4D_0\kappa)^{1/2} = \pm 1.189$ and a barrier at $x = 0$ with barrier height $F^\ddagger = F(0) - F(x_\pm) = D_0a^4 + \lambda^2a^2/2\kappa + \lambda^4/16D_0\kappa^2 \approx 10$.

A wide range of AFED parameters, $kT_x \in \{5...10\}$ and $m_x \in \{300...1000\}$, accurately reproduce the free energy profile. The set of parameters resulting in the closest agreement with the analytical profile with the greatest efficiency are $kT_x = 5$ and $m_x = 1000$. For comparison, d-AFED/TAMD simulations were performed with the fictitious coordinate s and momentum p_s subject to the adiabatic decoupling and high temperature. The chosen parameters are consistent with the original AFED simulations, i.e., $kT_s = \{5, 7.5\}$ and $m_s = 1000$. The fictitious coordinate s was then coupled to the system with a harmonic constant, $\kappa_s = 3000$, and the additional harmonic potential $V_{\text{coupling}}(x, s) = {}^1/_2\kappa_s(x - s)^2$.

Free energy profiles computed using the AFED (dashed lines) and the d-AFED/TAMD (solid and dot-dashed lines) approaches for simulations of 1×10^9 steps are depicted in Figure 1a. These profiles are nearly indistinguishable and accurately reconstruct the analytical free energy profile (dotted lines). It is only through close inspection of the minima (Figure 1b) and the barrier (Figure 1c) that minuscule deviations between the surfaces are revealed.

The convergence of the minima is computed by $\zeta_{\min}(t) = 0.5(|F_{x+}(t) - F_{x+,\text{an}}| + |F_{x-}(t) - F_{x-,\text{an}}|)$, where $F_{x\pm,\text{an}}$ are the analytical values of the free energy at the locations $x_\pm$ of the two minima, and the convergence of the whole surfaces is measured with $\zeta(t) = (1/N)\sum_{i=1}^{N}|F(i, t) - F_{\text{an}}(i)|$. Figure 2 shows that d-AFED/TAMD does not converge the surface as quickly as AFED but nevertheless does an excellent job generating the surface to within a small error (attributed to the finite value of the harmonic coupling, κ). The very slow convergence of $\zeta(t)$ to zero is due to small residual errors in the barrier, as Figure 1c illustrates. It is clear, however, from Figure 2 (bottom) that the d-AFED/TAMD approach quickly and accurately generates the location and relative energies of the minima.

Interestingly, the d-AFED/TAMD approach, while understandably unable to outperform its AFED predecessor in efficiency, is able to perform at nearly the same level of accuracy with only a slight degradation in efficiency.

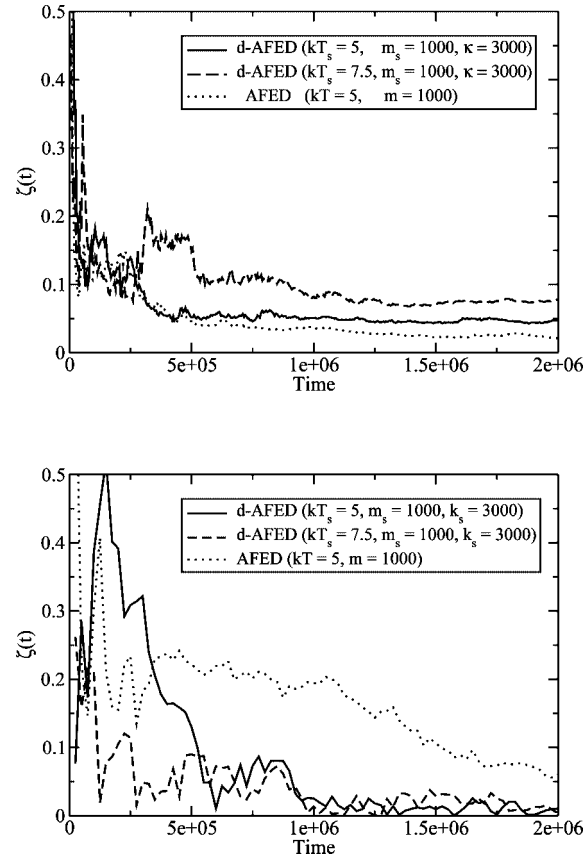

Figure 2. Convergence of the entire surface (top) and minima (bottom) for the double well model. AFED results in dotted lines, and d-AFED/TAMD results plotted in solid and dashed lines.

3.1. Collective Variables of Interest. The advantage of the d-AFED/TAMD scheme is that it simplifies the implementation of AFED at the cost of one additional parameter, the harmonic coupling κ. In addition, the new approach allows collective variables that were previously not possible within the original AFED method to be studied. In ref 23, a detailed transformation

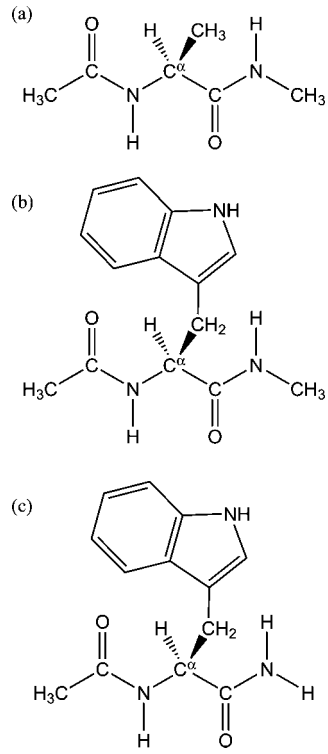

Figure 3. Schematics of the dipeptides: (a) alanine dipeptide, (b) NATMA (*N*-acetyl-tryptophan-methylamide), and (c) NATA (*N*-acetyl-tryptophan-amide).

needed to study dihedral angles within the AFED scheme was given. Using d-AFED/TAMD, that transformation is no longer required. Morevoer, other collective variables such as radius of gyration, R_G, and number of hydrogen bonds, N_H, for which no straightforward transformation exists, can now be studied with d-AFED/TAMD.

To demonstrate the ease and flexibility of d-AFED/TAMD, the studies performed in this paper will use these two aforementioned collective variables (CVs): R_G and N_H. The radius of gyration (R_G) of the heavy atoms of the backbone of a polypeptide is defined as

$$R_G = \sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}\left(\mathbf{r}_i - \frac{1}{N_b}\sum_{j=1}^{N_b}\mathbf{r}_j\right)^2} \tag{31}$$

where N_b is the number of heavy backbone atoms (N, C_α, C). This CV has been shown to be useful in studying the degree to which a polypeptide is folded or unfolded.[35] Similar to the end-to-end distance, R_{ee}, the radius of gyration distinguishes between proteins in the completely folded, completely unfolded, and globular disordered states.

A CV of interest, complementary to the radius of gyration, is the number of hydrogen bonds, N_H. The number of hydrogen bonds is a powerful way to discern the degree of order in proteins. Larger numbers of hydrogen bonds corresponds to a higher degree of order. This CV is defined as

$$N_H = \sum_{i=1}^{O}\sum_{j=1}^{H}\frac{1-\left(\frac{\mathbf{r}_i - \mathbf{r}_j}{d_0}\right)^6}{1-\left(\frac{\mathbf{r}_i - \mathbf{r}_j}{d_0}\right)^{12}} \tag{32}$$

where O is the number of oxygens, H is the number of hydrogens, and $d_0 = 2.5$ Å is the distance parameter. The proper parametrizations and choices for these collective variables is discussed further in the Appendix.

3.2. Simulation Details. Four systems were chosen for the present study: *N*-acetyl-Ala-methylamide (alanine dipeptide, Figure 3a), *N*-acetyl-tryptophan-methylamide (NATMA, Figure 3b), *N*-acetyl-tryptophan-amide (NATA, Figure 3c), and *N*-acetyl-$(Ala)_6$-methylamide (alanine hexamer, Figure 12). The following protocol was implemented in the subsequent studies. Initial configurations for the gas phase simulations were obtained by equilibrating each of the systems for an average of 500 ps−1 ns at system temperatures of $T = 300$ K (for alanine dipeptide) and $T = 423$ K (for NATMA and NATA).

The solution phase simulations of the alanine dipeptide and the alanine hexamer were performed in cubic boxes with periodic boundary conditions. Each box contains one solute molecule and in a bath of TIP3P waters. Each box was equilibrated at constant volume (NVT), then at constant pressure (NPT) for 200 ps, and again at constant volume (NVT) for another 500 ps.

The electrostatics were treated via the smooth particle-mesh Ewald method,[36] and the short-range forces were switched off at a distance of 12 Å. The dipeptides were parametrized with the CHARMM22,[37,38] AMBER95,[39,40] and the all-atom version of the OPLS (OPLS-AA)[41,42] force fields as indicated, and the TIP3P[43] water model was used as a solvent. The alanine hexamer was parametrized with the AMBER force field (with the parm94 parameters). The SHAKE[44] and RATTLE[45] algorithms were used to constrain the water OH bonds to a tolerance of 10^{-6}. The NVT distribution in each case was generated by coupling a generalized Gaussian moment thermostat (GGMT)[30,46] to each degree of freedom in the system. The equations of motion were integrated using the multiple time scale algorithm, r-RESPA,[32,34] to exploit the various separations of time scales. Typically, 3−5 r-RESPA steps per time step were used. The integrator is derived using the propagator factorization schemes in refs 34 and 30. Recently, Ezra introduced a rigorously time-reversible integrator for the GGMT method.[47] Although we are interested in testing this algorithm in the present context, we have not yet been able to make it efficient enough for our purposes. Production runs were performed up to 20 ns, with an innermost r-RESPA time step of 0.5 fs. All calculations presented in this paper were performed with the PINY_MD code.[48] All one-dimensional free energy profiles were processed with an 11-point Savitsky−Golay filter[49] while all two-dimensional surfaces were processed by five-point filters applied across each slice.

4. Results and Discussion: Alanine Dipeptide

The alanine dipeptide,[50−64] one of the most well-studied systems, is depicted in Figure 3a. In applying the d-AFED/TAMD method with the CHARMM22 force field, the two collective variables of interest chosen were the radius of gyration (R_G) and the number of intramolecular hydrogen bonds (N_H). Corresponding coordinates s_1 and s_2 and their conjugate momenta p_{s_1} and p_{s_2} were added, and these coordinates were treated as the slow variables with masses $m_{s_1,s_2} = 10m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 300$ K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 5.4×10^6 K Å^{-2} and 5.4×10^6 K, respectively. Parts a and b of Figure 4 show the two-

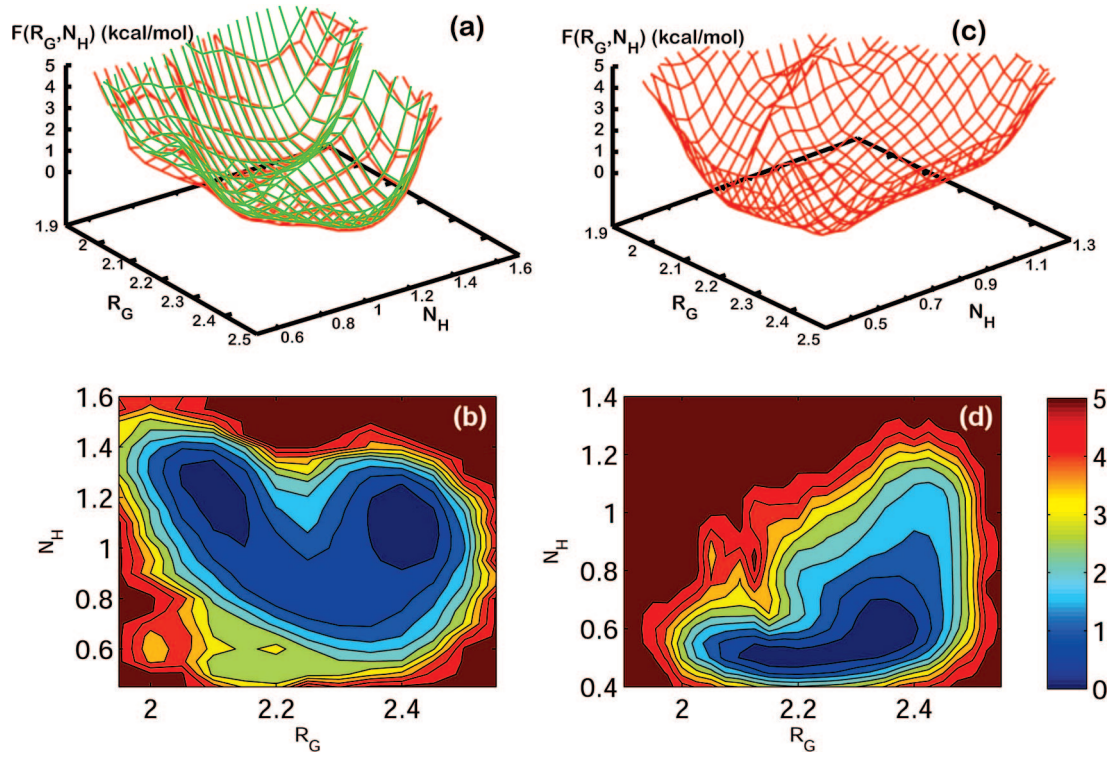

Figure 4. (a) Free energy surface $F(R_G,N_H)$ for the alanine dipeptide in the gas phase generated using d-AFED/TAMD for 1 ns (red) and traditional MD for 12 ns (green). (b) Contour map corresponding to the free energy surface in a for the alanine dipeptide in the gas phase. (c) Free energy surface $F(R_G,N_H)$ for the alanine dipeptide in solution generated using the d-AFED/TAMD for 1 ns. (d) Same as in (b), except corresponding to free energy surface in (c), i.e., solvated alanine dipeptide.

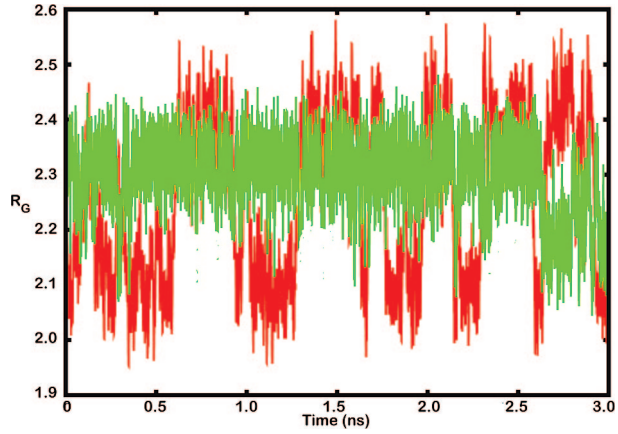

Figure 5. Trajectories of the traditional MD simulation (green) and d-AFED/TAMD simulation (red) in the R_G dimension for the alanine dipeptide in the gas phase.

dimensional free energy surface $F(R_G,N_H)$ and the contour map, respectively, generated with the d-AFED/TAMD method for alanine dipeptide in the gas phase (total simulation length = 1 ns). Figure 4a also depicts the same free energy surface computed with a long-time traditional MD simulation (total simulation length = 12 ns).[65] This figure shows that the d-AFED/TAMD approach and MD are in relatively good agreement, but as expected the traditional MD surface has not adequately sampled the higher energy regions of the configurational space on this time scale. Figure 5 clearly shows that the traditional MD trajectory clearly undergoes far fewer barrier-crossing events than the d-AFED/TAMD trajectory, even in the case of the relatively low-barrier alanine dipeptide free energy surface.

The characteristics of the free energy surface generated in Figure 4a,b are consistent with the other studies performed on the alanine dipeptide with the CHARMM22 force field.[23,64,52,56,57,63] Specifically, the three characteristic minima, {C5, $C7_{eq}$, α_R }, are found. The $C7_{eq}$ extended conformation is located at $(R_G, N_H) = (2.1, 1.25)$ and is the global minimum. The second extended conformation, C5, is located at $(R_G, N_H) = (2.4, 1.1)$ and is $\sim$0.15 kcal mol^{-1} above the $C7_{eq}$ conformation. The final, and most controversial, minimum is the right-handed helical (α_R) conformation. Located at $(R_G, N_H) = (2.15, 0.5)$, α_R is clearly a local minimum with relative free energy $\sim$2.75 kcal mol^{-1} above $C7_{eq}$. In the studies of the Ramachandran surfaces, it is unclear as to whether α_R is truly a minimum or simply a shoulder on the surface. Using this set of collective variables, it is clear that α_R is indeed a minimum (Table 1). Additionally, it should be noted that the surfaces, $F(R_G, N_H)$, tend to be significantly less rough than the Ramachandran surfaces. Consequently, the temperatures used to drive the fictitious coordinates need not be as high. Even though the temperature of the extended variables is the same as that of the physical variables, the driving action of the extended variables is sufficient to produce the enhanced sampling revealed in Figure 5.

The solution phase simulations were performed by solvating the alanine dipeptide in 216 TIP3P water molecules in a 18.8 Å cubic periodic box. In the same manner as the gas phase calculations, two additional degrees of freedom (s_1, s_2) were added and were treated as the slow variables with masses $m_{s_1,s_2} = 75m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 300$ K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 1.0×10^7 K Å^{-2} and 1.0×10^7 K, respectively.

Parts c and d of Figure 4 show the two-dimensional free nergy surface $F(R_G, N_H)$ and the contour map, respectively, generated with the d-AFED/TAMD method for alanine dipeptide in solution (total simulation length = 1 ns). The characteristic

TABLE 1: Relative Free Energies (kcal/mol) of Minima on the Free Energy Surface, $F(R_G, N_H)$, of Alanine Dipeptide with d-AFED and Compared to Selected values from the Literature for the Free Energy Surface, $F(\phi, \psi)$

	gas phase				solution			
	C5	$C7_{eq}$	α_R	$C7_{ax}$	α_R	β	C5	$C7_{ax}$
d-AFED	0.15	0.0	2.75	*b*	0.0	0.15	1.75	4.25
AFED	0.44[a]	0.0	3.35[a]	2.8[a]	0.0	0.2[c]	0.2[c]	4.5
US		0.0		2.7	0.0	0.3[c]	0.3[c]	3.85
MetaDyn	0.24	0.0	3.18	1.15	0.0	1.0[c]	1.0[c]	4.8
(R_G, N_H)	(2.4, 1.1)	(2.1, 1.25)	(2.15, 0.5)		(2.15, 0.5)	(2.35, 0.6)	(2.4, 1.1)	(2.05, 0.9)

[a] Adjusted results from ref 23. [b] $C7_{eq}$ and $C7_{ax}$ overlap in this CV space and are indistinguishable. [c] Free energy surfaces computed with ϕ and ψ as collective variables cannot distinguish between β, C5, and $C7_{eq}$ minima here.

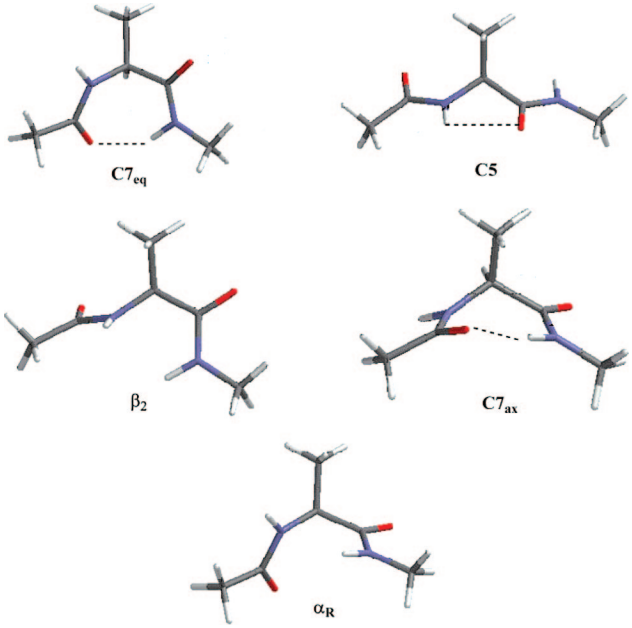

Figure 6. C5, $C7_{eq}$, $C7_{ax}$, β, and α_R conformations of alanine dipeptide.

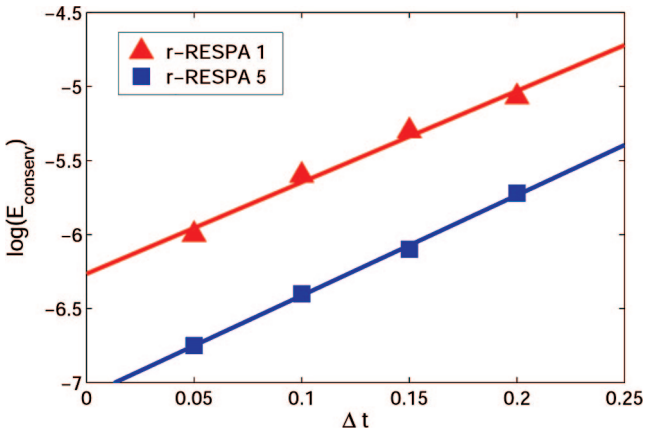

Figure 7. Log of energy conservation vs timestep for simulations without (red line) and with (blue line) r-RESPA factorization on the coupling/driving potential. The value of 5 for r-RESPA indicates a factor of 5 between the large and small time steps. A value of 1 is equivalent to no r-RESPA on this term. Note that r-RESPA is still used in both cases to treat the bonds and bends, torsions, and nonbonded forces with different time steps.

minima for this surface are the three extended conformations, {β, C5, $C7_{ax}$}, and the right-handed helical conformation α_R. Figure 4d shows that, unlike the case of the alanine dipeptide in gas phase, the right-handed helical α_R conformation is the global minimum in solution and occurs at $(R_G, N_H) = (2.15, 0.5)$. (For a detailed discussion on the values of N_H, see the Appendix.) The extended β and $C7_{ax}$ conformations occur at $(R_G, N_H) = (2.35, 0.6)$ and (2.05,0.9), respectively, with relative energies 0.15 and 4.25 kcal mol^{-1}, respectively. The figure also reveals that there is a significant contribution of the extended C5 conformation, $(R_G, N_H) = (2.4, 1.1)$, with relative free energy 1.75 kcal mol^{-1} above the α_R conformation. These results are all in good agreement with previous simulations published in the literature[23,64,52,56,57,63] (Table 1), many of which require longer simulation times (10−100 ns).

The most significant difference is that unlike the Ramachandran surfaces, where the three extended conformations are indistinguishable, we observe three distinct minima using R_G and N_H as CVs. (Figure 6 depicts the five most important minima for the alanine dipeptide in gas phase and solution.)

Finally, as discussed above, the r-RESPA multiple time scale factorization[32,34] leads to significant increases in performance without the need to decrease the size of the fundamental time step. All of the simulations performed used the r-RESPA factorization on the intramolecular interactions (as well as a secondary factorization for the torsions). Additionally, simulations of alanine dipeptide were performed with an r-RESPA level for the d-AFED/TAMD coupling potentials. In the notation of refs 32 and 34 the Liouville operator is decomposed into the following terms:

$$iL = iL_{nb} + iL_{tors} + iL_{intra} + iL_{d\text{-}AFED} \tag{33}$$

where iL_{nb} and iL_{tors} correspond to the nonbonded and torsional forces, respectively, iL_{intra} corresponds to the bond and bend forces, and iL_{d-AFED} corresponds to the d-AFED/TAMD harmonic coupling term, which is the last term in eq 21. These operators are defined as

$$iL_{nb} = \sum_{i=1}^{N} \mathbf{F}_{nb,i} \cdot \frac{\partial}{\partial \mathbf{p}_i}$$

$$iL_{tors} = \sum_{i=1}^{N} \mathbf{F}_{tors,i} \cdot \frac{\partial}{\partial \mathbf{p}_i}$$

$$iL_{intra} = \sum_{i=1}^{N} \mathbf{F}_{intra,i} \cdot \frac{\partial}{\partial \mathbf{p}_i}$$

$$iL_{d\text{-}AFED} = \sum_{i=1}^{N} \left[\frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{r}_i} + \mathbf{F}_{d\text{-}AFED,i} \cdot \frac{\partial}{\partial \mathbf{p}_i} \right] \tag{34}$$

Here, $\mathbf{F}_{nb,i}$, $\mathbf{F}_{tors,i}$, $\mathbf{F}_{intra,i}$, and $\mathbf{F}_{d\text{-}AFED,i}$ are the nonbonded, torsional, intramolecular, and d-AFED/TAMD harmonic cou-

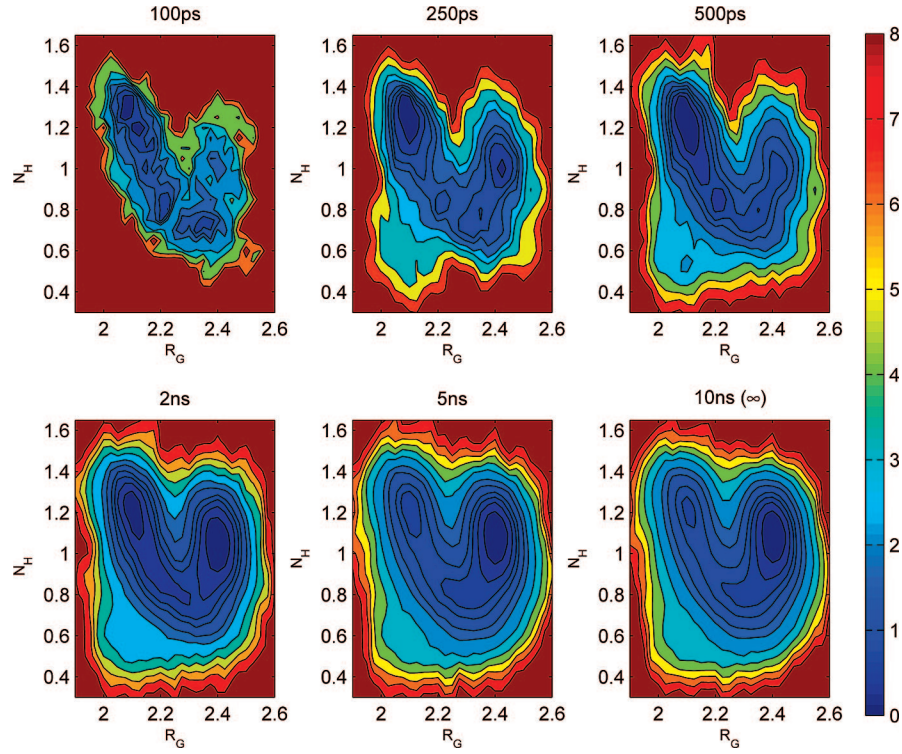

Figure 8. Evolution over time of the free energy surface, $F(R_G,N_H)$, for *N*-acetyl-tryptophan-methylamide (NATMA) in the gas phase with the CHARMM22 force field.

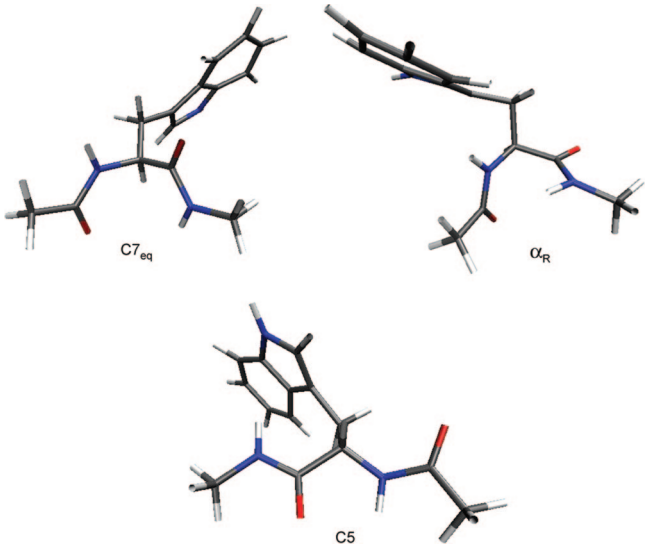

Figure 9. $C7_{eq}$, α_R, and C5 conformations of alanine dipeptide.

pling forces on particle *i*. The single time-step propagator $\exp(iL\Delta t)$ is then factorized according to the scheme

$$e^{iL\Delta t} \approx e^{iL_{nb}\Delta t}\{e^{iL_{tors}\delta t_1}[e^{iL_{intra}\delta t_2}(e^{iL_{d\text{-}AFED}\delta t_3})^{n_3}e^{iL_{intra}\delta t_2}]^{n_2}e^{iL_{tors}\delta t_1}\}^{n_1}e^{iL_{nb}\Delta t} \quad (35)$$

where the times steps Δt, δt_1, δt_2, and δt_3 satisfy the relations $\Delta t = n_1\delta t_1 = n_1n_2\delta t_2 = n_1n_2n_3\delta t_3$. The central operator $\exp(iL_{d-AFED}\delta t_3)$ is further factorized into the standard velocity Verlet integrator by applying the Trotter as follows:

It should also be noted that the thermostat variables were evolved using XO-RESPA.[34] That is, the thermostats were updated before and after the evolution of the physical system and not inside.[34] The effect of using r-RESPA on the size of the time step accessible is shown in Figure 7. The figure depicts only the effect of r-RESPA on the harmonic coupling term $(\kappa/2)\sum_\alpha(q_\alpha(\mathbf{r}) - s_\alpha)^2$. In both simulations r-RESPA is used to integrate bonds and bends, torsions, and nonbonded forces with different time steps;[34] specifically, three steps for bonds and bends are taken for each torsion step, and five torsion steps are taken for each nonbonded step. The energy conservation is plotted versus time step for simulation with (blue line) and without (red line) r-RESPA on the d-AFED/TAMD harmonic coupling term. It is clear from the results that significantly larger timesteps can be achieved when using r-RESPA and without significant increase in the CPU time.

5. Results and Discussion: *N*-acetyl-tryptophan-methylamide

N-acetyl-tryptophan-methylamide (NATMA)[66−71] is depicted in Figure 3b. The reason for choosing this as an example system is that the bulkier side chain from the tryptophan gives rise to roughness in the energy landscape that is more characteristic of that encountered in protein folding. In applying the d-AFED/TAMD approach, the two collective variables of interest chosen were the radius of gyration (R_G) and the number of intramolecular hydrogen bonds (N_H). As with the alanine dipeptide, corresponding coordinates, s_1 and s_2, were added, and these coordinates were treated as the slow variables with masses $m_{s_1,s_2} = 150m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 450$ K, slightly higher than the temperature of the physical variables. Here, s_1 and s_2 are coupled to R_G and N_H with harmonic coupling constants 5.4×10^6 K Å^{-2} and 5.4×10^6 K, respectively. Figure 8 shows the evolution

TABLE 2: Locations and Relative Free Energies (kcal/mol) of Minima on the Free Energy Surface, $F(R_G, N_H)$, of NATMA with d-AFED Using Three MM Force Fields

CHARMM22			AMBER95			OPLS-AA		
C5	(2.40, 1.08)	0.0	$C7_{eq}$	(2.10, 1.27)	0.0	C5	(2.45, 1.08)	0.0
$C7_{eq}$	(2.11, 1.18)	1.0	C5	(2.44, 1.01)	0.75	$C7_{eq}$	(2.10, 1.35)	1.25
α_R	(2.11, 0.60)	3.2	α_R[a]	(2.12, 0.70)	0.75	α_R		
$F^{\ddagger}_{C5/C7}$	(2.17, 0.94)	0.9	$F^{\ddagger}_{C5/C7}$	(2.29, 0.77)	1.9	$F^{\ddagger}_{C5/C7}$	(2.31, 0.70)	3.6

[a] The $C7_{eq}$ and α_R minima are connected.

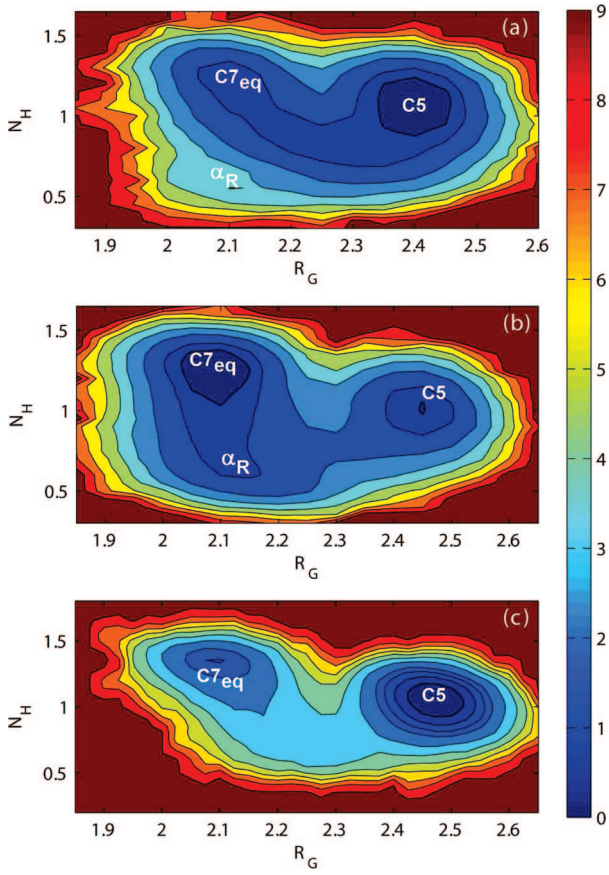

Figure 10. Free energy surfaces, $F(R_G, N_H)$, computed for *N*-acetyl-tryptophan-methylamide (NATMA) in the gas phase with the (a) CHARMM22, (b) AMBER95, and (c) OPLS-AA force fields.

of the free energy surface for NATMA in gas phase over time, as computed using the CHARMM22 force field.

The figure clearly shows that the general shape of the surface is established in as little as 100 ps. After only 250 ps of total simulation time, the three minima (C5, $C7_{eq}$, α_R) are already clearly discernible. The surface has been completely converged after only 4 ns of total simulation time. The global minimum is the extended C5 conformation located at $(R_G, N_H) = (2.4, 1.08)$. The next minimum of importance is the $C7_{eq}$ extended conformation with relative free energy 1 kcal mol^{-1} above the minimum and located at $(R_G, N_H) = (2.11, 1.18)$. The final minimum is the right-handed helical conformation (α_R) located at $(R_G, N_H) = (2.11, 0.6)$ with relative free energy of +3.2 1 kcal mol^{-1}. The three minimum free energy conformations of NATMA in gas phase using the CHARMM22 force field are depicted in Figure 9.

The same calculations, with the same d-AFED conditions, were also performed using the two other most popular force fields: AMBER95[39,40] and OPLS-AA.[41,42] The results for the locations and minima of the three force fields with NATMA are given in Table 2, and the corresponding free energy surfaces are displayed in Figure 10. Examining the free energy surfaces in Figure 10, we see that all three force fields agree that the extended conformations (C5 and $C7_{eq}$) are the preferred conformations for NATMA in gas phase. This is the extent to which all three of the force fields agree.

The first significant disagreement between the results generated with the three force fields is the lack of an α_R contribution in the OPLS-AA free energy surface (Figure 10c). In in the free energy surfaces computed with both CHARMM22 and AMBER95 (Figure 10, part a and b, respectively), a minimum is found in the right-handed helical region of the surfaces with relative free energy +3.2 and +0.75 kcal mol^{-1}, respectively. The free energy surface generated with OPLS-AA shows no α_R character whatsoever. Another discrepancy in the results is that while the free energy surfaces generated by both CHARMM22 and OPLS-AA agree that the C5 conformation is the preferred conformation, it appears that in AMBER95 that the $C7_{eq}$ (and α_R) conformation is lower in energy (with C5 having relative free energy at +0.75 kcal mol^{-1}). Finally, the barrier to isomerization from C5 to $C7_{eq}$, and the difference in the relative energies of these minima, is significantly higher in the OPLS-AA simulation than in either of the other two simulations. It is also interesting to note that the barriers to isomerization in NATMA in gas phase (0.9−3.6 kcal mol^{-1}) are all higher, as predicted, than the barrier to isomerization of alanine dipeptide in gas phase (0.5 kcal mol^{-1}), as predicted.

Experimental results for the conformational preferences of NATMA in the gas phase,[67] studied by UV and IR spectroscopy of the conformers, indicates that the lowest energy conformation of NATMA in gas phase (at $T = 423$ K) is the C5 extended conformation. The next lowest energy structure was a $C7_{eq}$ extended conformer with energy +0.73 kcal mol^{-1} above the C5 minimum with barrier to isomerization at 0.93 kcal mol^{-1}. The results obtained from the CHARMM22 simulation most closely reproduce the experimental results. The simulations with the OPLS-AA force field correctly identify the lowest energy structure but underestimate the probability of the $C7_{eq}$ structure and overestimate the height of the barrier between C5 and $C7_{eq}$. The presence or lack of an α_R minimum for NATMA in gas phase is not discussed or presented in the literature.

6. Results and Discussion: *N*-acetyl-tryptophan-amide

N-acetyl-tryptophan-amide (NATA) is depicted in Figure 3c. The primary interest in this compound is to contrast its conformational preferences with the "capped" tryptophan amide, NATMA. Most computational studies of small peptide systems use *N*-acetyl caps at the N-terminus and methylamide caps at the C-terminus of the polypeptides. The rationale behind adding these additional backbone-like segments is that their inclusion leads to behavior more consistent with that of bulk proteins. The comparison of the conformation preferences of NATMA and NATA will shed some light on the necessity of this practice.

Simulation parameters identical to those used for the NATMA simulations were used in the NATA simulations. The coordi-

TABLE 3: Locations and Relative Free Energies (kcal/mol) of Minima on the Free Energy Surface, $F(R_G, N_H)$, of NATA with d-AFED Using Three MM Force Fields

CHARMM22			AMBER95			OPLS-AA		
C5	(2.05, 1.58)	0.0	$C7_{eq}$	(1.85, 1.85)	0.0	C5	(2.12, 1.6)	0.0
$C7_{eq}$	(1.83, 1.80)	0.6	C5	(2.1, 1.52)	1.0	$C7_{eq}$	(1.84, 2.03)	1.0
α_R	(1.81, 1.14)	3.0	α_R	(1.74, 0.90)	0.0	α_R[b]	(1.79, 1.25)	6.75
$F^{\ddagger}_{C5/C7}$	(1.9, 1.4)	0.6[a]	$F^{\ddagger}_{C5/C7}$	(2.0, 1.38)	2.0	$F^{\ddagger}_{C5/C7}$	(1.95, 1.4)	3.75

[a] The barrier plateaus into the $C7_{eq}$ minimum. [b] This is more of a shoulder than a minimum.

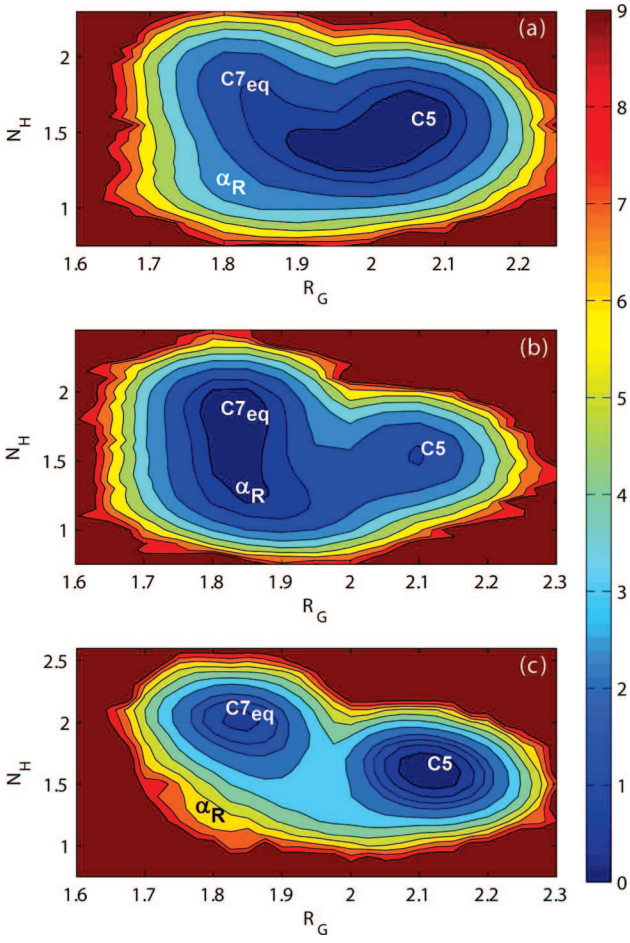

Figure 11. Free energy surfaces, $F(R_G,N_H)$, computed for *N*-acetyl-tryptophan-amide (NATA) in the gas phase with the (a) CHARMM22, (b) AMBER95, and (c) OPLS-AA force fields.

TABLE 4: Change in Relative Free Energy, in kcal mol^{-1}, from Capped (NATMA) to Uncapped (NATA) Tryptophan Dipeptide

CHARMM22		AMBER95		OPLS-AA	
C5	0.0	$C7_{eq}$	0.0	C5	0.0
$C7_{eq}$	−0.4	C5	+0.25	$C7_{eq}$	−0.25
α_R	−0.2	α_R	−0.75	α_R	
$F^{\ddagger}_{C5/C7}$	−0.3	$F^{\ddagger}_{C5/C7}$	+0.1	$F^{\ddagger}_{C5/C7}$	+0.15

nates, s_1 and s_2, were added and were treated as the slow variables with masses $m_{s_1,s_2} = 150m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 450$ K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 5.4×10^6 K Å^{-2} and 5.4×10^6 K, respectively. Table 3 details the results for the conformational preferences of NATA in gas phase evaluated using d-AFED/TAMD and the three force fields (CHARMM22, AMBER95, and OPLS-AA).

Comparing the free energy surfaces for NATMA from Figure 10a−c to those computed for NATA, in Figure 11a−c, it is clear that the overall conformational landscape is not significantly affected. The extended conformations remain the dominant configurations. The most significant differences observed are the emergence of right-handed helical conformations in the OPLS-AA and AMBER95 free energy surfaces (Figure 11, part b and c, respectively). Additionally, we see that the addition of a second H on the N-terminus, effectively doubling the number of atoms able to form a seven-membered C7 ring, increases the fraction of $C7_{eq}$ conformations at equilibrium in all three simulations. Table 4 summarizes the effect of capping on the conformational preferences of these tryptophan dipeptides.

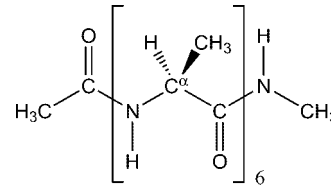

Figure 12. Schematics of the alanine hexamer (*N*-acetyl-(alanine)$_6$-methylamide).

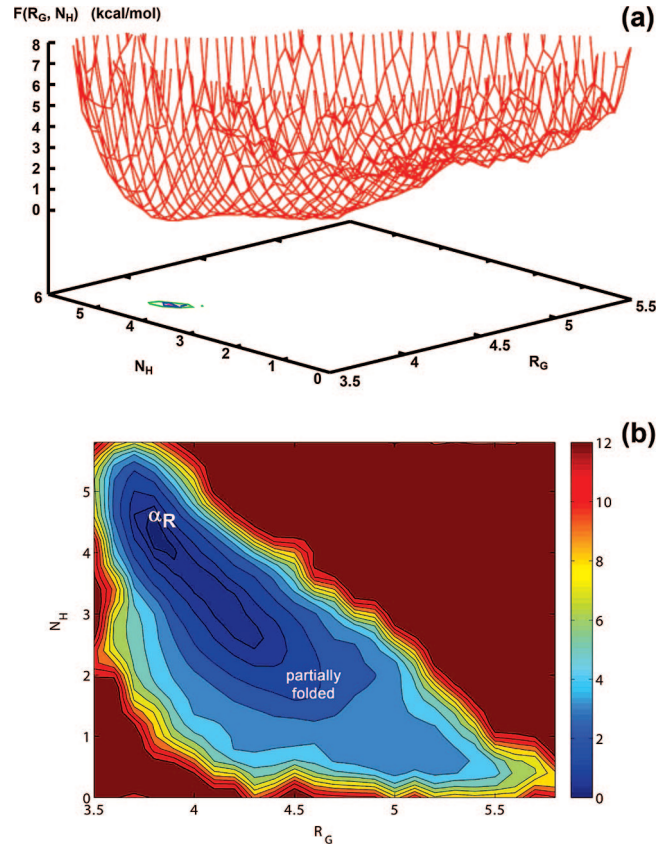

Figure 13. (a) Free energy surface $F(R_G,N_H)$ and (b) contour plot computed for *N*-acetyl-$(Ala)_6$-methylamide (alanine hexamer) in solution with the AMBER force field.

$$\exp[iL_{\text{d-AFED}}\delta t_3] = \exp\left[\frac{\delta t_3}{2}\sum_{i=1}^{N}\mathbf{F}_{\text{d-AFED},i}\cdot\frac{\partial}{\partial \mathbf{p}_i}\right]\times \exp\left[\delta t_3\sum_{i=1}^{N}\frac{\mathbf{p}_i}{m_i}\frac{\partial}{\partial \mathbf{r}_i}\right]\exp\left[\frac{\delta t_3}{2}\sum_{i=1}^{N}\mathbf{F}_{\text{d-AFED},i}\cdot\frac{\partial}{\partial \mathbf{p}_i}\right] + \mathcal{O}(\delta t_3{}^3) \quad (36)$$

7. Results and Discussion: Alanine Hexamer

The alanine hexamer (*N*-acetyl-(Ala)$_6$-methylamide) is depicted in Figure 12. It is known that this six-residue peptide exhibits helical properties in solution.[72] Furthermore, computational studies of the alanine hexamer in solution, parametrized with the AMBER force field, results in helical conformational minima.[73]

The simulation was performed by solvating the alanine hexamer in 698 TIP3P water molecules in a 27.9737 Å cubic periodic box. The molecule was started in its completely extended conformation and then equilibrated for 200 ps at constant volume (NVT), 1 ns at constant pressure (NPT), and finally 500 ps at constant volume (NVT).

In applying the d-AFED/TAMD method with the AMBER (parm94) force field, the two collective variables of interest chosen were the radius of gyration (R_G) and the number of intramolecular hydrogen bonds (N_H). Corresponding coordinates, s_1 and s_2, were added, and these coordinates were treated as the slow variables with masses $m_{s_1,s_2} = 15m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} =$ 600 K, while the physical variables were kept at a temperature of 300 K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 5.4 × 10^6 K Å^{-2} and 5.4 × 10^6 K, respectively.

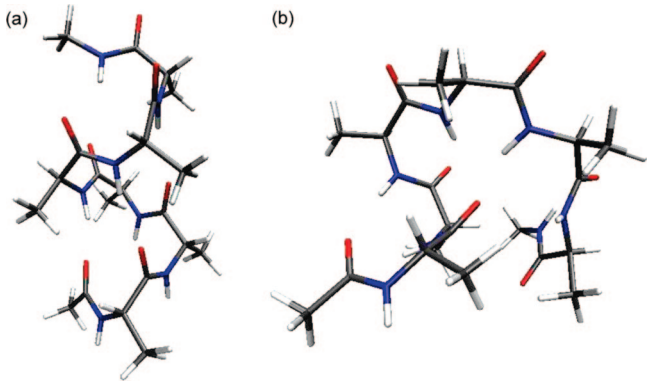

Figure 14. (a) Right-handed helical conformation of the alanine hexamer (global minimum). (b) Misfolded, extended, conformation of the alanine hexamer. All waters have been deleted for clarity.

TABLE 5: H-Bond Contributions to the N_H Collective Variable (NATMA in Gas Phase with CHARMM22)

C5			C7$_{eq}$		
interaction	$\mathbf{r}_{OH}$ (Å)	n_H	interaction	$\mathbf{r}_{OH}$ (Å)	n_H
HNCO	3.1	2 × 0.2	HNCO	3.15	2 × 0.2
C5	2.33	0.67	C7	2.07	0.8
total:		$N_H = 1.1$	total:		$N_H = 1.2$
α_R			**β**		
interaction	$\mathbf{r}_{OH}$ (Å)	n_H	interaction	$\mathbf{r}_{OH}$ (Å)	n_H
HNCO	3.15	2 × 0.2	HNCO	3.17	0.2
			H-bond	3.20	2 × 0.19
total:		$N_H = 0.4$	total:		$N_H = 0.58$

The free energy surface computed from a relatively short production run of 5 ns, for the alanine hexamer in solution, leads to the free energy surface shown in Figure 13. The most significant feature of this surface is the presence of a global minimum at (R_G, N_H) = (3.8, 4.4), corresponding to a right-handed helical, α_R, conformation (Figure 14a). Furthermore, there is also the presence of a large region of deformed/partial helices all with free energies lower than +2 1 kcal mol^{-1} above the global minimum. These preliminary results are very consistent with previous results[73] for the alanine hexamer using the AMBER force field.

8. Conclusion

The adiabatic free energy dynamics (AFED)[21−23,25] is a powerful approach for the efficient generation of multidimensional free energy surfaces in a set of collective variables of particular interest in a system. AFED uses an adiabatic decoupling of these collective variables of interest from the remaining degrees of freedom and then subjects these variables only to a high temperature. The principal disadvantage of the method is the need to transform the coordinate space into a set of generalized coordinates that explicitly contain these collective variables. The improvement by Maragliano and Vanden-Eijnden,[24] called temperature-accelerated molecular dynamics (TAMD), circumvents the need for explicit coordinate transformations by introducing a set of extended phase space variables, similar to the metadynamics approach[20] and then subjects these extended phase space variables to the adiabatic decoupling and enhanced temperature. TAMD, therefore, increases the flexibility of AFED and renders it considerably easier to implement in existing molecular dynamics codes.

The ability of AFED and TAMD to generate free energy profiles and surfaces efficiently depends on the thermostating method employed to maintain the two temperatures under adiabatic conditions. In this paper, we have shown that a formulation of TAMD that employs the generalized Gaussian moment thermostat (GGMT) together with multiple time scale (r-RESPA) integration leads to a method that clearly is able to generate free energy surfaces for complex systems *directly from the dynamics*. It is important to emphasize that the use of r-RESPA allows the value of κ to be increased, thereby ensuring better convergence to the correct free energy surface, without sacrificing efficiency. Of course, this time scale separation can only be pushed so far before well-known resonance effects[75,76] limit the large time step. We expect that the use of resonant-free multiple time step integrators, such as that proposed by Minary et al.[77] (or a version of this method that incorporates the GGMT thermostat) will alleviate this problem, and future work will involve applying such resonant-free techniques to d-AFED simulations.

The new TAMD formulation, which we call driven AFED or d-AFED was used to compute conformational free energy surfaces, $F(R_G,N_H)$, in the radius of gyration R_G and number of hydrogen bonds N_H for the alanine dipeptide in gas phase and solution, *N*-acetyl-tryptophan-methylamide (NATMA) in gas phase, and *N*-acetyl-tryptophan-amide (NATA) in gas phase. The results indicate that d-AFED/TAMD approaches the efficiency and accuracy of the original AFED without the inconvenience of the coordinate transformations. For the alanine dipeptide, the conformational preferences computed using d-AFED/TAMD agree very well with the literature values and even go so far as to answer questions surrounding the previously ambiguous α_R gas-phase minimum. NATMA and NATA, on the other hand, have not been studied nearly as extensively as

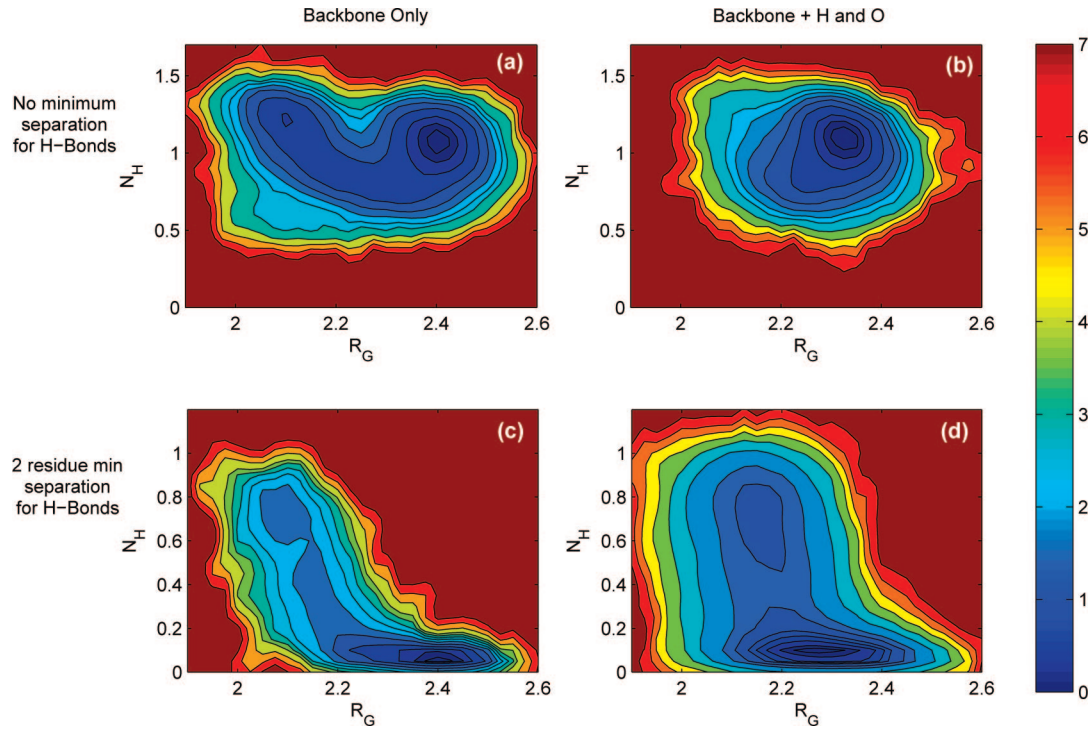

Figure 15. Free energy surfaces, $F(R_G,N_H)$, computed for *N*-acetyl-tryptophan-methylamide (NATMA) in the gas phase with the CHARMM22 for a variety of CV parameters. (a, c) Only backbone C and N are included in the calculation of R_G (b and d include all non-side-chain atoms). (c, d) A minimum separation of two residues between atom involved in hydrogen bonding is imposed when calculating N_H (a and b have no restrictions placed on N_H).

the alanine dipeptide. The results are, however, in good agreement with the experimental data that are available. From the results, it appears that CHARMM22 and OPLS-AA are the force fields of choice for these types of calculations. While it is clear that the three force fields agree on the basics of the conformational preferences of NATMA, there are significant differences for this system as well. We expect the new d-AFED/TAMD approach to be a powerful and flexible one for computing free energy surfaces. Future work will involve extending the method to ab initio molecular dynamics for exploring free energies of chemical bond-breaking and -forming processes in the gas phase, solution, and biomolecules.

Acknowledgment. The authors thank Eric Vanden-Eijnden for useful discussions. This work was supported by NSF CHE-0704036 and PRF 45485-AC5. J.B.A. acknowledges the Margaret and Herman Sokol Graduate Fellowship.

Appendix. Collective Variable Analysis

The definitions of the two collective variables used in this study, the number of hydrogen bonds and the radius of gyration, N_H and R_G, respectively, are not unique. The purpose of this appendix, therefore, is to analyze briefly the information encoded in the different definitions and examine how the choice of the collective variable definitions affects the free energy surfaces. While the radius of gyration collective variable is clearly defined in eq 31, there exists some ambiguity as to which "backbone" atoms should be included. One approach is to include all atoms that are not side-chain atoms: all of the C, N, O, H. A second approach is to only include the actual backbone C and N atoms that make up the framework of the backbone.

Regarding the N_H collective variable, the most significant factor rests in the option of whether or not to impose a minimum residue separation between O and H involved in H-bonding. Including H-bonds separated by at least two residues includes C7-type interactions but excludes dipeptide-bond (HNCO) and C5 interactions. Reducing the separation imposed to one residue allows for the inclusion of HNCO interactions. Without an imposed H-bond separation, any and all H−O interactions are included. (In each of these cases, only H's on N or O atoms are included.) Because the overall N_H collective variable is a sum over a potentially large number of intramolecular interactions, it is reasonable to expect that the values not be limited to discrete integer values. Table 5 shows standard H-bond interaction contributions to the major conformers' N_H values.

To fully illustrate the significance of the specific choice of collective variables, identical simulations were performed using two choices each for R_G and N_H. The NATMA dipeptide system with the CHARMM22 force field was selected, and the simulations were performed with the same parameters as in section 6. Figure 15 shows the free energy surfaces generated for this 2 × 2 matrix of CVs. Inspection of the surfaces shows that there is not visible α_R minimum when all non-side-chain atoms are included in the calculation of R_G (right column). It is also clear that the spacing between the $C7_{eq}$ (less extended) and C5 (more extended) minima decreases to the point where there is almost no visible minimum in $C7_{eq}$ (Figure 15b). The lack of separation of the C5 and $C7_{eq}$ minima is rectified by imposing a minimum separation of two residues between H-bonds (Figure 15d), but the α_R minimum is still not isolated. The choice of CVs that leads to the free energy surface with the most information, and clearly defined minima, is Figure 15a. Here, no separation between residues is imposed for the H-bond interaction in N_H and only the C and N atoms of the backbone are included in the calculation of R_G.

References and Notes

(1) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.
(2) Torrie, G. M.; Valleau, J. P. *J. Comput. Chem.* **1977**, *23*, 187.
(3) Berne, B. J., Ed. *Modern Theoretical Chemistry V*; Plenum: New York, 1977.
(4) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(5) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(6) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737.
(7) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607.
(8) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.
(9) Tesi, M.; Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155.
(10) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749.
(11) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703.
(12) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13711.
(13) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.
(14) Voter, A. F. *Phys. Rev. B* **1998**, *57*, R13985.
(15) Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050.
(16) Wang, F. G.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101.
(17) Siepmann, J. I.; Frenkel, D. *Mol. Phys.* **1992**, *75*, 59.
(18) Zhu, Z.; Tuckerman, M. E.; Samuelson, S. O.; Martyna, G. J. *Phys. Rev. Lett.* **2002**, *88*, 100201.
(19) Minary, P.; Tuckerman, M. E.; Martyna, G. J. *SIAM J. Sci. Comput.* **2008**, *30*, 2055.
(20) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
(21) Rosso, L.; Tuckerman, M. E. *Mol. Simul.* **2002**, *28*, 91.
(22) Rosso, L.; Minary, P.; Zhu, Z.; Tuckerman, M. E. *J. Chem. Phys.* **2002**, *116*, 4389.
(23) Rosso, L.; Abrams, J. B.; Tuckerman, M. E. *J. Phys. Chem. B* **2005**, *109*, 4162–4167.
(24) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168.
(25) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 074115.
(26) The technique was first presented at a CECAM meeting on free energy calculations held in 2000 from June 19−21. The proceedings of this meeting were published in *Molecular Simulation*, Vol. 28, Issues 1−2 (2002), which contains ref 21.
(27) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(28) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533.
(29) Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110.
(30) Liu, Y.; Tuckerman, M. E. *J. Chem. Phys.* **2000**, *112*, 1685–1700.
(31) Samuelson, S.; Martyna, G. J. *J. Chim. Phys.* **1997**, *94*, 1503.
(32) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(33) Butkov, E. *Mathematical Physics*; Addison-Wesley: Reading, MA, 1968.
(34) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117.
(35) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435.
(36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(37) Brooks, B.; Bruccoleri, R.; Olafson, D.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(38) MacKerel Jr. , A.; Brooks III, C.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M.; John Wiley & Sons: Chichester, 1998; Vol. 1 of *The Encyclopedia of Computational Chemistry*; pages 271−277.
(39) Pearlman, D.; Case, D.; Caldwell, J.; Rossand, W.; T. E.; Cheatham, I.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
(40) Case, D.; T. E.; Cheatham, I.; Darden, T.; Gohlke, H.; Luo, R., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(41) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(42) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(45) Andersen, H. *J. Comput. Phys.* **1983**, *52*, 24.
(46) It has been demonstrated in previous works that the GGMT thermostats are more reliable than the Nosé−Hoover chains[74] for systems subject to the nonequilibrium adiabatic conditions employed in these types of simulations (where high-energy barriers are involved). Therefore, the use of the GGMT thermostating approach, in particular, the coupling of each degree of freedom to a GGMT thermostat, is recommended.
(47) Ezra, G. S. *J. Chem. Phys.* **2006**, *125*, 034104.
(48) Tuckerman, M. E.; Yarne, D. A.; Samuelson, S. O.; Martyna, G. J. *Comput. Phys. Commun.* **2000**, *128*, 333.
(49) Press, W. H.; Teukolsky, S. A.; Vetterling, W. Y.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: Carmbridge, 1992.
(50) Mezei, M. *J. Comput. Phys* **1987**, *68*, 1420–1426.
(51) Roterman, I. K.; Lamberg, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Mol. Biol. Struct. Dyn.* **1989**, *7*, 421.
(52) Tobias, D. J.; C. L.; Brooks, I. *J. Phys. Chem.* **1992**, *96*, 3864–3870.
(53) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
(54) Olender, R.; Elber, R. *J. Chem. Phys.* **1996**, *105*, 9299–9315.
(55) Samuelson, S.; Hughes, A.; Martyna, G. J. *J. Chim. Phys* **1997**, *94*, 1503.
(56) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568–5579.
(57) Apostolakis, J.; Ferrara, P.; Caflish, A. *J. Chem. Phys.* **1999**, *110*, 2099–2108.
(58) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 3213.
(59) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451.
(60) Feig, M.; MacKerell, A. D.; C. L.; Brooks, I. *J. Phys. Chem. B* **2003**, *107*, 2831.
(61) Weise, C. F.; Weisshaar, J. C. *J. Phys. Chem. B* **2003**, *107*, 3265–3277.
(62) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–1716.
(63) A. D. Mackerell, J.; Feig, M.; C. L.; Brooks, I. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(64) Ensing, B.; de Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73–81.
(65) It has already been shown that the original AFED method accurately and rigorously computes these surfaces. In ref 23 AFED was benchmarked against umbrella sampling with WHAM.
(66) Dian, B. C.; Longarte, A.; Zwier, T. S. *Science* **2002**, *296*, 2369.
(67) Dian, B. C.; Longarte, A.; Mercier, S.; Evans, D. A.; Wales, D. J.; Zwier, T. S. *J. Chem. Phys.* **2002**, *117*, 10688.
(68) Dian, B. C.; Longarte, A.; Winter, P. R.; Zwier, T. S. *J. Chem. Phys.* **2004**, *120*, 133.
(69) Dian, B. C.; Florio, G. M.; Clarkson, J. R.; Longarte, A.; Zwier, T. S. *J. Chem. Phys.* **2004**, *120*, 9033.
(70) Yurtsever, E.; Yuret, D.; Erman, B. *J. Phys. Chem. A* **2006**, *110*, 13933.
(71) Bagchi, S.; Kim, Y. S.; Charnley, A. K.; A. B.; Smith, I.; Hochstrasser, R. M. *J. Phys. Chem. B* **2007**, *111*, 3010.
(72) Kennedy, R. J.; Tsang, K. Y.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 934.
(73) Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. *Chem. Phys. Lett.* **2005**, *401*, 312.
(74) Martyna, G. J.; Tuckerman, M. E.; Klein, M. L. *J. Chem. Phys.* **1992**, *97*, 2635.
(75) Schlick, T.; Mandziuk, M.; Skeel, R. D.; Srinivas, K. *J. Comput. Phys.* **1998**, *140*, 1.
(76) Ma, Q.; Izaguirre, J. A.; Skeel, R. D. *SIAM J. Sci. Comput.* **2003**, *24*, 1951.
(77) Minary, P.; Tuckerman, M. E.; Martyna, G. J. *Phys. Rev. Lett.* **2004**, *93*, 150201.

JP805039U