

Temperature-Accelerated Sampling and Amplified Collective Motion with Adiabatic Reweighting to Obtain Canonical Distributions and Ensemble Averages

Yue Hu,^{†,‡} Wei Hong,^{†,‡} Yunyu Shi,^{†,‡,§} and Haiyan Liu^{*,†,‡,§}

[†]School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China

[‡]Key Laboratory of Structural Biology, Chinese Academy of Science, Hefei, Anhui 230027, China

[§]Hefei National Laboratory for Physical Sciences at Microscale, Hefei, Anhui 230027, China

ABSTRACT: In molecular simulations, accelerated sampling can be achieved efficiently by raising the temperature of a small number of coordinates. For collective coordinates, the temperature-accelerated molecular dynamics method or TAMD has been previously proposed, in which the system is extended by introducing virtual variables that are coupled to these coordinates and simulated at higher temperatures (Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2005**, *426*, 168–175). In such accelerated simulations, steady state or equilibrium distributions may exist but deviate from the canonical Boltzmann one. We show that by assuming adiabatic decoupling between the subsystems simulated at different temperatures, correct canonical distributions and ensemble averages can be obtained through reweighting. The method makes use of the low-dimensional free energy surfaces that are estimated as Gaussian mixture probability densities through maximum likelihood and expectation maximization. Previously, we proposed the amplified collective motion method or ACM. The method employs the coarse-grained elastic network model or ANM to extract collective coordinates for accelerated sampling. Here, we combine the ideas of ACM and of TAMD to develop a general technique that can achieve canonical sampling through reweighting under the adiabatic approximation. To test the validity and accuracy of adiabatic reweighting, first we consider a single *n*-butane molecule in a canonical stochastic heat bath. Then, we use explicitly solvated alanine dipeptide and GB1 peptide as model systems to demonstrate the proposed approaches. With alanine dipeptide, it is shown that sampling can be accelerated by more than an order of magnitude with TAMD while correct distributions and canonical ensemble averages can be recovered, necessarily through adiabatic reweighting. For the GB1 peptide, the conformational distribution sampled by ACM-TAMD, after adiabatic reweighting, suggested that a normal simulation suffered significantly from insufficient sampling and that the reweighted ACM-TAMD distribution may present significant improvements over the normal simulation in representing the local conformational ensemble around the folded structure of GB1.

INTRODUCTION

Molecular dynamics (MD) or Monte Carlo (MC) simulations are widely applied to study the conformational distributions and dynamics of biomolecules in solution.^{1–4} Conventional constant temperature MD or MC simulations sample conformations or configurations from a canonical Boltzmann distribution.^{1,5} Averaging over the sampled configurations may give various properties of interest of the system. This approach requires the simulations to be long enough so that relevant parts of the conformational space can be sampled sufficiently.

This requirement cannot always be fulfilled, often due to slow transitions in the configurational space between different low energy regions that are separated by barriers associated with high energies. In the canonical Boltzmann distribution, the probability associated with a configuration of Cartesian coordinates $\mathbf{r}^N = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ is proportional to the Boltzmann factor $\exp[-V(\mathbf{r}^N)/k_B T]$, where N is the number of particles, $V(\mathbf{r}^N)$ is the potential energy, k_B is the Boltzmann constant, and T is the temperature. Because of the form of the Boltzmann factor, barrier regions associated with high potential energies are sampled with low probabilities, leading to rare barrier crossing or slow conformational transitions.

To deal with this difficulty, there is constant interest in accelerated sampling methods, namely, methods to achieve more extensive sampling in the conformational space at the same computational costs as conventional simulations. To this end, a range of methods have been proposed and continuously improved over the past decades (see, for examples, reviews 6 and 7 and references therein). Many of the proposed methods share the common idea of trying to increase the probability of sampling in the barrier regions based on the basic form of the Boltzmann distribution.

According to this form, increased sampling in the barrier regions can be achieved either by changing the potential energy function so that the energies of configurations in these regions relative to those in the minimum regions are lowered or by increasing the temperature so that regions associated with higher energies are sampled with larger probabilities.

Both strategies are conceptually simple. However, applying them in a simple-minded manner can rarely be very useful. This is because changing either the potential energy or the

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: January 27, 2012

Published: April 13, 2012



temperature causes the distribution of conformations to deviate from the original canonical distribution. If the deviation is too large, most of the sampled configuration will no longer be physically relevant or meaningful; thus no effective acceleration of sampling can be achieved.^{8,8} For example, to simulate all of the degrees of freedom of the system at a too high temperature is in general not a good idea for accelerated sampling, especially for systems with explicit solvent. To do so leads to indiscriminative expansion of the accessible conformational space into high energy regions, most part of which are inaccessible to the system at the ambient temperature and contribute little to the canonical distribution at normal temperature. One way to get around this problem is to target not all but only a small subset of degrees of freedom for accelerated sampling. In what follows, we will focus on this type of method, because it is especially promising for the efficient sampling of conformational distributions of proteins, including large amplitude motions related to their functions.⁴ It has been shown that after proper choices of variables, such motions can often be described by slow but significant changes in only a small subset of (collective) degrees of freedom, while the dynamics of the remaining degrees of freedom can be treated as relatively small and fast fluctuations.

For the purpose of accelerated sampling along a small number of conformational coordinates, methods to modify the potential energies with biasing potentials that depend on these coordinates have been proposed. These methods include conformational flooding,⁹ local elevation,¹⁰ and meta-dynamics.¹¹ A common feature of them is that the biasing potentials are constructed adaptively to lower the transition barriers, differentiating these methods from approaches that use prespecified restraints to bias the sampling toward specific regions in the conformational space.¹² In the meta-dynamics approach proposed by Laio and Parrinello¹¹ and widely applied,¹³ the system phase space is extended by introducing a virtual system that moves on the free energy surface or potential of mean forces along chosen collective conformational variables. The biasing potential acts on the virtual system and is built up gradually to flatten out the free energy surface, offering the possibility to estimate the free energy surface by an inversion of sign of the biasing potential. Nevertheless, it should be mentioned that the performance (e.g., efficiency, convergence rates, and errors) of the various adaptive biasing methods depends critically on parameters of the respective algorithms and to determine suitable parameters for a given problem is a highly sophisticated and delicate process.^{6,9–11,13}

To avoid the time-consuming and parameter-sensitive adaptive construction of the biasing potential, accelerated sampling can be achieved by increasing the temperature associated with specific degrees of freedom. For example, to accelerate the search of optimum ligand positions relative to a receptor in molecular docking, Di Nola et al. first suggested to simulate the center of mass motions of the ligand at a higher temperature while simulating the remaining motions of the system at the ambient temperature.^{14,15} In 2003, we proposed the amplified collective motion (ACM) method,^{16,17} in which a few slowest collective modes of a protein are derived from a coarse-grained elastic network model.^{18,19} Then, motions in the subspace spanned by these modes are coupled to a higher temperature bath in atomic simulations. The original ACM scheme works via projecting and recombining the atomic momenta and does not aim at estimating any canonical distributions or averages.¹⁶

More recently, significant progress has been made through the temperature-accelerated MD (TAMD) approach proposed by Maragliano and Vanden-Eijnden.²⁰ In this approach, virtual variables are introduced as in meta-dynamics. Unlike in meta-dynamics, the virtual variables in TAMD are simply simulated at a higher temperature with a stochastic algorithm. It has been noted²¹ that TAMD is related to another temperature-based accelerated sampling method called adiabatic molecular dynamics.^{22,23} If the virtual variables in TAMD are associated with sufficiently large masses, the so-called adiabatic decoupling condition can be met. Then, the extended system as a whole may reach an equilibrium distribution, in which the distribution of physical degrees of freedom that are orthogonal to the accelerated coordinates will remain canonical at normal temperature, while the distribution of the virtual variables is a high-temperature Boltzmann one on a free energy surface generated by the orthogonal degrees of freedom at normal temperature.^{20,22,23} As results of the high-temperature distribution of the virtual variables, the accelerated conformational coordinates associated with the virtual variables follow the high-temperature but not the normal-temperature distribution.⁸ Thus, all degrees of freedom considered, conformations sampled through adiabatic decoupling, including TAMD, do not follow the canonical distribution.⁸

It is highly desirable to establish a scheme to recover the canonical Boltzmann distributions and to obtain correct thermodynamic ensemble averages in adiabatic decoupling simulations such as TAMD. This will allow ensemble averages and distributions of various physical properties to be investigated using accelerated simulations. In ref 8 on adiabatic decoupling, a reweighting factor was given. It depends on an ensemble average involving the interaction energy between the adiabatically decoupled heavy and light degrees of freedoms as an exponential factor. Because of the large fluctuations of this energy, this formula was found to be not practically useful.⁸ We note that for meta-dynamics, the so-called well-tempered meta-dynamics approach has been proposed in order to reach a converging biasing potential and a converging distribution.^{24,25} The converged distribution is also not canonical, and a reweighting scheme has been proposed.²⁶

The present work has been inspired by previous studies of TAMD,²⁰ well-tempered meta-dynamics,^{24,26} adiabatic decoupling,^{22,23} and our own ACM approach.^{16,17} In the current study, we investigate whether and how configurations sampled in adiabatic decoupling simulations can be reweighted in a way that is theoretically correct and practically useful. The proposed approach needs the adiabatic free energy surface along the accelerated coordinates or variables; thus we will call it adiabatic reweighting. In this study, this free energy surface is derived from the probability density estimated as Gaussian mixtures through maximum likelihood and expectation maximization.²⁷ We also investigate the combination of ACM, which targets collective coordination from a coarse-grained elastic network model for temperature-based accelerated sampling, and TAMD, which uses extending variables and allows adiabatic reweighting to be applied. We investigate whether this ACM-TAMD approach can accelerate explorations in the conformational space and, more importantly, whether ACM-TAMD with adiabatic reweighting can improve the sampling of canonical conformational distributions, a fundamental goal of molecular simulation. We consider three test systems. One is *n*-butane in the gas phase, for which the results of adiabatic reweighting can be compared with accurate canonical distributions. The other

two are the alanine dipeptide and the GB1 peptide,²⁸ both in explicit water solvent.

THEORY AND METHOD

1. Extended Phase Space Simulations for Accelerated Sampling.

Following ref 20, we consider M collective conformational variables $\{q_\alpha(\mathbf{r}^N), \alpha = 1, \dots, M\}$ defined using coordinates \mathbf{r}^N in the physical configurational space. Associated with each $q_\alpha(\mathbf{r}^N)$, a virtual variable Q_α can be defined. We then consider an extended system with configurational coordinates $\{\mathbf{r}^N, \mathbf{Q}^M\} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, Q_1, Q_2, \dots, Q_M\}$ and the following total potential energy function:²⁰

$$V_{\text{extended}}(\mathbf{r}^N, \mathbf{Q}^M) = V(\mathbf{r}^N) + \sum_{\alpha=1}^M \frac{1}{2} K_\alpha [Q_\alpha - q_\alpha(\mathbf{r}^N)]^2 \quad (1)$$

in which K_α 's are restraining force constants that should be chosen large enough so that a collective variable $q_\alpha(\mathbf{r}^N)$ cannot deviate too far away from the respective virtual variable Q_α . Besides the interactions in eq 1, each of the virtual variables Q_α can be assigned a fictitious mass m_α . To sample configurations in the extended phase space, a standard constant temperature (the temperature noted as T_l) molecular dynamics simulation can be applied to the physical coordinates, and stochastic simulation at a higher temperature (noted as T_h) can be applied to the virtual variables. More specifically, the time evolution of the virtual variables can be simulated by the following Langevin equations of motion:

$$\frac{dQ_\alpha}{dt} = v_\alpha \quad (2)$$

and

$$\frac{dv_\alpha}{dt} = \frac{1}{m_\alpha} \left(-\frac{\partial V_{\text{extended}}}{\partial Q_\alpha} + R_\alpha \right) - \gamma_\alpha v_\alpha \quad (3)$$

in which R_α is the random force and γ_α is the friction coefficient. The random force obeys a one-dimensional Gaussian probability distribution with a zero mean and a variance of $\langle R_\alpha^2 \rangle = 2m_\alpha \gamma_\alpha k_B T_h$. We use the leapfrog stochastic dynamics algorithm²⁹ to integrate eqs 2 and 3. We assume that the virtual variables have the unit of length. So in eq 1, the virtual variables implicitly carry a proportional constant with a proper unit and a value of 1 to make the difference $Q_\alpha - q_\alpha(\mathbf{r}^N)$ meaningful. An alternative view is to assume that Q_α has the same unit as $q_\alpha(\mathbf{r}^N)$, but the mass m_α takes a proper unit so that $1/2m_\alpha(dQ_\alpha/dt)^2$ has the unit of energy. The two views do not make any differences in practice.

If the masses m_α are large enough, the virtual variables and the physical system can be assumed to be adiabatically decoupled. Under this assumption, we may first consider the configurational partition function of the physical coordinates for given virtual \mathbf{Q}^M :^{8,22,23}

$$Z_{\text{physical}}(\mathbf{Q}^M) = \int \exp(-1/(k_B T_l)[V(\mathbf{r}^N) + \sum_{\alpha=1}^M \frac{1}{2} K_\alpha (Q_\alpha - q_\alpha(\mathbf{r}^N))^2]) d\mathbf{r}^N \quad (4)$$

The configurational partition function for the extended virtual variables is given by

$$Z_{\text{virtual}} = \int \exp(-1/k_B T_h F_{\text{physical}}(\mathbf{Q}^M)) d\mathbf{Q}^M \quad (5)$$

with

$$F_{\text{physical}}(\mathbf{Q}^M) = -k_B T_l \ln Z_{\text{physical}}(\mathbf{Q}^M) \quad (6)$$

Note the different temperatures in eqs 4–6, and adiabatic decoupling has been assumed in these equations. In the limit of very large K_α 's, the effective potential $F_{\text{physical}}(\mathbf{Q}^M)$ acting on the virtual variables is a good approximation to the true physical free energy surface or potential of mean forces at T_l along the respective collective coordinates,²⁰ namely,

$$\begin{aligned} F_{\text{physical}}(\mathbf{Q}^M) &\approx F(\mathbf{Q}^M) \\ &= -k_B T_l \ln \int \exp(-1/(k_B T_l) V(\mathbf{r}^N)) \\ &\quad \prod_{\alpha=1}^M \delta(Q_\alpha - q_\alpha(\mathbf{r}^N)) d\mathbf{r}^N \end{aligned} \quad (7)$$

in which δ is the Dirac δ function. In further discussions, we will no longer discriminate between $F_{\text{physical}}(\mathbf{Q}^M)$ and $F(\mathbf{Q}^M)$. We note that the approximation made in eq 7 should not be confused with the adiabatic decoupling approximation that has already been assumed in eqs 4–6.

Because the distribution of the virtual variables is governed by the partition function in eq 5, the potential of mean forces $F_{\text{physical}}(\mathbf{Q}^M)$ or $F(\mathbf{Q}^M)$ can be reconstructed from the equilibrium probability density distribution of \mathbf{Q}^M or $\rho(\mathbf{Q}^M)$:

$$F(\mathbf{Q}^M) = -k_B T_h \ln \rho(\mathbf{Q}^M) \quad (8)$$

Note that the virtual variables \mathbf{Q}^M and the corresponding collective conformational coordinates of the physical systems are sampled according to the high temperature distribution. If slow motions along these coordinates comprise the major sampling bottlenecks in usual MD, accelerated sampling can be achieved.

2. Collective Variables from the Coarse-Grained Elastic Network Model of Proteins. For proteins, the residue-based elastic network model or ANM (anisotropic network model)^{18,19} can be applied to obtain an “essential subspace”³⁰ for slow motions, namely, a subspace spanned by the few lowest frequency collective coordinates obtained from the following coarse-grained harmonic potential energy function:¹⁹

$$V_{\text{ANM}}(\mathbf{r}^N) = \sum_{i,j} \frac{1}{2} k_{ij}^{\text{ANM}} (r_{ij} - r_{ij}^{\text{ref}})^2 \quad (9)$$

in which the summation is over pairs of amino acid residues; r_{ij} and r_{ij}^{ref} are inter-residue distances in the current configuration \mathbf{r}^N and in a reference configuration $\mathbf{r}_{\text{ref}}^N$ respectively. The reference configuration is the minimum energy configuration for V_{ANM} . In this work we use the positions of C^α atoms to represent the positions of amino acid residues and use the “pre-equilibrated” starting configuration for accelerated sampling as the reference configuration. k_{ij}^{ANM} is given by

$$k_{ij}^{\text{ANM}} = \begin{cases} 1, & \text{if } r_{ij}^{\text{ref}} \leq r_{\text{cutoff}}^{\text{ANM}} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Here, appropriate units are assumed for various quantities in eqs 9 and 10. The $r_{\text{cutoff}}^{\text{ANM}}$ is the single cutoff parameter for the ANM model. Other choices of distance-dependent k_{ij}^{ANM} (see

for example, ref 31) may have been used but may not affect the results much, as we are only concerned with the slowest few modes and only the directions of them.

By applying standard normal-mode analysis, the few slowest normal modes for V_{ANM} can be obtained as eigenvectors associated with the few smallest eigenvalues of the Hessian matrix of V_{ANM} (after excluding the six zero eigenvalues associated with overall translations and rotations).^{18,19} The Hessian matrix is the matrix of second order derivatives of V_{ANM} with respect to coordinates evaluated at $\mathbf{r}_{\text{ref}}^N$. We denote the eigenvector associated with the μ th smallest eigenvalue as $\mathbf{e}_\mu = (e_{\mu 1}, e_{\mu 2}, \dots, e_{\mu 3N})^t$. For configuration \mathbf{r}^N , the collective coordinates along this eigenvector are given by

$$q_\mu(\mathbf{r}^N) = (\mathbf{r}^N - \mathbf{r}_{\text{ref}}^N) \mathbf{e}_\mu \quad (11)$$

These coordinates can be used as the collective variables $q_\alpha(\mathbf{r}^N)$ in eqs 1–8, and we will call this approach ACM-TAMD. When \mathbf{r}^N is sampled by simulations, the $\mathbf{r}^N - \mathbf{r}_{\text{ref}}^N$ in eq 11 may include overall translations and rotations. This is not a problem for the calculation of $q_\mu(\mathbf{r}^N)$ or equivalently $q_\alpha(\mathbf{r}^N)$, but the translations and rotations do change the forces on the physical system due to interactions with the virtual coordinates. Thus, at every time step in ACM-TAMD, $\mathbf{r}_{\text{ref}}^N$ is translated and rotated “on the fly” to minimize its root-mean-square deviations (RMSD) of C^α positions from \mathbf{r}^N , and the eigenvector \mathbf{e}_μ 's are transformed accordingly. The transformed $\mathbf{r}_{\text{ref}}^N$ and \mathbf{e}_μ 's are used in eq 11.

3. Adiabatic Reweighting. Through adiabatic TAMD, the equilibrium configurational distribution follows different canonical distributions in the subspace spanned by the collective variables and in the subspace orthogonal to these variables. The former is at temperature T_h and the latter at temperature T_l . Thus, the physical system as a whole deviates from its canonical Boltzmann distribution, and direct averaging over the sampled configurations will not give physically meaningful ensemble averages. For any configuration \mathbf{r}^N , the biasing factor in adiabatic TAMD sampling is

$$\begin{aligned} \frac{\rho_{\text{canonical}}(\mathbf{r}^N)}{\rho_{\text{ad-TAMD}}(\mathbf{r}^N)} &= \frac{\exp(-1/(k_B T_l) F(\mathbf{q}(\mathbf{r}^N)))}{\exp(-1/(k_B T_h) F(\mathbf{q}(\mathbf{r}^N)))} \times \text{Const} \\ &= \exp((\frac{1}{k_B T_h} - \frac{1}{k_B T_l}) F(\mathbf{q}(\mathbf{r}^N))) \times \text{Const} \end{aligned} \quad (12)$$

in which $\rho_{\text{ad-TAMD}}$ stands for configurational probability density for adiabatic TAMD with temperatures T_h and T_l , and $\rho_{\text{canonical}}$ stands for the canonical Boltzmann distribution with temperature T_b , $\mathbf{q}(\mathbf{r}^N) = \{q_\alpha(\mathbf{r}^N), \alpha = 1, \dots, M\}$, and Const is a constant independent of \mathbf{r}^N for proper normalization. Thus, for any quantity $O(\mathbf{r}^N)$, its canonical ensemble average can be obtained as

$$\begin{aligned} \langle O \rangle_{\text{canonical}} &= \int O(\mathbf{r}^N) \rho_{\text{canonical}}(\mathbf{r}^N) d\mathbf{r}^N \\ &= \int O(\mathbf{r}^N) \rho_{\text{ad-TAMD}}(\mathbf{r}^N) \exp((\frac{1}{k_B T_h} - \frac{1}{k_B T_l}) \\ &\quad F(\mathbf{q}(\mathbf{r}^N))) \times \text{Const} d\mathbf{r}^N \\ &= \left\langle O \exp((\frac{1}{k_B T_h} - \frac{1}{k_B T_l}) F(\mathbf{q}(\mathbf{r}^N))) \right\rangle_{\text{ad-TAMD}} \\ &\quad \times \text{Const} \end{aligned} \quad (13)$$

Const is determined through normalization, i.e., when averaging over L configurations sampled through adiabatic TAMD:

$$\text{Const} = [\sum_{l=1}^L \exp((\frac{1}{k_B T_h} - \frac{1}{k_B T_l}) F(\mathbf{q}(\mathbf{r}_l^N)))]^{-1} \quad (14)$$

Because $(1/k_B T_h) - (1/k_B T_l) < 0$, configurations in lower free energy regions will receive larger weights. This contributes to the practical usefulness of the reweighting formulas 13 and 14, because the low free energy regions are likely to be well-sampled and corresponding free energies accurately determined. Thus, adiabatic reweighting is not prone to inaccuracies of the free energy surface in high free energy regions, as these regions associated with large statistical errors have smaller weights in the final averages. In fact, statistical uncertainties of adiabatically reweighted averages may even decrease from that of the unweighted averages (see results).

We note that eq 8, to estimate the free energy surface, and eq 13, to estimate canonical averages, have been derived on the basis of the assumption of adiabatic decoupling. Because the total potential energy in eq 1 must contain the coupling interactions $\sum_{\alpha=1}^M (1/2) K_\alpha [Q_\alpha - q_\alpha(\mathbf{r}^N)]^2$, exact or absolute adiabatic decoupling is only possible when the virtual masses in eq 3 are infinite. In practice, the virtual masses cannot be chosen to be infinite; there is always heat flowing from the virtual into the physical subsystem. However, when the virtual subsystem contains much fewer degrees of freedom than the physical subsystem, which is usually the case, the total amount of transferred heat can be negligibly small relative to the canonical energy fluctuations of the physical subsystem. Of course, the virtual masses need to be chosen large enough so that the rate of heat transfer is slow relative to the internal thermo-equilibration within the physical subsystem. A sufficient condition for this is that the “temperatures” of $q_\alpha(\mathbf{r}^N)$'s, the degree of freedom that directly accepts the heat from the virtual subsystem, are not significantly increased due to the heat transfer. This condition is overstringent for adiabatic reweighting to be a good approximation, because the actual requirement is not that the low temperature canonical distribution of $q_\alpha(\mathbf{r}^N)$ is maintained but that the canonical distributions of the orthogonal degrees of freedoms that receive heat from $q_\alpha(\mathbf{r}^N)$ are closely maintained. However, the condition can be easily checked, for example, by looking at the average of each of the $(1/2) K_\alpha [Q_\alpha - q_\alpha(\mathbf{r}^N)]^2$ term, for which the ideal value in the case of exact adiabatic decoupling and constant $F(\mathbf{Q}^M)$ (so the oscillation of q_α around Q_α is ideally harmonic) would be $(1/2) K_\alpha T_b$ and a value close to $(1/2) K_\alpha T_l$ but far from $(1/2) K_\alpha T_h$ should indicate adiabatic reweighting to be a good approximation.

To determine the weights in eq 13, one needs the free energy surface $F(\mathbf{Q}^M)$, which in turn needs to be determined from the probability density $\rho(\mathbf{Q}^M)$ according to eq 8. For low dimensional $\rho(\mathbf{Q}^M)$, it can be estimated using, e.g., a maximum-likelihood density estimator. For example, we use a weighted sum of n multivariate Gaussian functions to approximate $\rho(\mathbf{Q}^M)$,²⁷

$$\rho(\mathbf{Q}^M) \approx \tilde{\rho}(\mathbf{Q}^M) = \sum_{i=1}^n \alpha_i G_{\mu_i, \Sigma_i}(\mathbf{Q}^M) \quad (15)$$

in which $\tilde{\rho}(\mathbf{Q}^M)$ is the estimated distribution, α_i is the weight for the i th Gaussian function, G_{μ_i, Σ_i} is the i th multivariate

Gaussian function with mean μ_i and covariance Σ_i (Σ_i is not assumed to be diagonal). Given L sample points $\{\mathbf{Q}_l^M, l = 1, 2, \dots, L\}$ in the \mathbf{Q}^M space, the likelihood function of the parameter set $\theta = \{\alpha_i, \mu_i, \Sigma_i, i = 1, \dots, n\}$ can be written as

$$L(\theta|\{\mathbf{Q}_l^M, l = 1, 2, \dots, L\}) = \prod_{l=1}^L \sum_{i=1}^n \alpha_i G_{\mu_i, \Sigma_i}(\mathbf{Q}_l^M) \quad (16)$$

For a prespecified number of Gaussian functions n , this likelihood can be maximized using expectation maximization²⁷ to obtain optimum estimations of $\theta = \{\alpha_i, \mu_i, \Sigma_i, i = 1, \dots, n\}$. In principle, the optimum number of Gaussian functions can be determined by dividing the sampled points into a training set and a test set and determining a number n that leads to the largest likelihood for the test set using parameters estimated from the training set. For our example systems, we did not find this necessary because variation of the number of Gaussians in a reasonable range for each particular system makes little difference in final results.

4. Test Systems and Simulation Procedures.

***n*-Butane.** The model of *n*-butane ($\text{CH}_3\text{—CH}_2\text{—CH}_2\text{—CH}_3$) comprises four united atoms described by the GROMOS43A1³² force field. It was simulated in the gas phase using stochastic Langevin dynamics (SD).²⁹ The small size of the system allows the canonical distributions of the various geometric parameters to be determined accurately. To determine these distributions, a normal SD simulation of only the physical system was carried out with a stochastic bath temperature of 300 K and a uniform atomic frictional coefficient of 1.5 ps^{-1} . Then, two sets of TAMD simulations were performed on an extended system that contained, besides the physical subsystem, a virtual variable associated with the torsional angle around the central C–C bond. The force constant K_α that restrained the difference between the virtual variable and the actual torsional angle was set to be 10 000 kJ/mol/rad². While the physical subsystem was simulated by SD at 300 K, the virtual variable was simulated by SD with a bath temperature of 1600 K and a frictional coefficient of 1.0 ps^{-1} . We have carried out two separate sets of TAMD simulations, one with a virtual mass of 1 and the other with a virtual mass of 5 (the values are in atom mass units). The purpose of considering different virtual masses is to investigate the effects of imperfect adiabatic decoupling on the accuracy of adiabatic reweighting. For each value of the virtual mass, we have separately considered different atomic frictional coefficients (0.5 ps^{-1} , 1.5 ps^{-1} , and 2.5 ps^{-1}) for the physical subsystem. The purpose of using different frictional coefficients is to look at the effects of different rates of thermo-equilibration within the physical subsystem.

Alanine Dipeptide. This system consists of one alanine dipeptide (alanine capped with CH_3CO at the N terminus and with NHCH_3 at the C terminus) described by the GROMOS43A1³² force field in a cubic box containing 1288 explicit SPC³³ waters. A 1 ns constant pressure simulation at 300 K yielded an average box side length of 3.41733 nm. The box side length has been fixed at this value in subsequent constant volume simulations.

Using this small system, we aim at investigating the effects of different parameters in TAMD on sampling acceleration, on estimating free energy surfaces, and on obtaining canonical ensemble averages with adiabatic reweighting. Thus, we have performed a normal 30 ns constant temperature simulation at 300 K as well as a series of TAMD simulations that targeted the

peptide backbone torsional angles φ and ψ as collective variables for accelerated sampling. In the different TAMD simulations, the virtual variables have been assigned different masses (10, 50, 200, or 1000 in atom mass unit), coupled to baths of different temperatures (800 or 1600 K) using stochastic simulations with different friction coefficients (0.1, 0.5, or 1.0 in ps^{-1}). See also Tables 1 and 2 for the different sets

Table 1. Summary of Results^a of the TAMD Simulations of *n*-Butane

m_{virtual}	$\gamma_{\text{physical}} (\text{ps}^{-1})$	$T_{\text{virtual}} (\text{K})$	$T_{\text{physical}} (\text{K})$	$\langle V_{\text{coupling}} \rangle (\text{kJ/mol})$
1	0.5	1460	321.8	1.77
	1.5	1380	313.3	1.47
	2.5	1333	312.0	1.39
	5	1557	301.8	1.35
	0.5	1541	301.0	1.29
	2.5	1518	302.3	1.28

^aFrom left to right, data in the columns are the mass of the virtual variable, the atomic frictional coefficient for the physical subsystem, the average temperature of the virtual variable, the average temperature of the physical subsystem, and the average coupling interaction energy between the virtual and the physical subsystems.

of parameters compared. The restraining force constant K_α is 1000 kJ/mol/rad². K_α should to be large enough to restrain $|Q_\alpha - q_\alpha(\mathbf{r}^N)|$, which has the order of magnitude of $(k_B T_l / K_\alpha)^{1/2}$ under the adiabatic assumption, and small enough to allow the 2 fs time step used in the simulations (as K_α increases, the oscillation frequency of $|Q_\alpha - q_\alpha(\mathbf{r}^N)|$ also increases). On the basis of short simulations of a few picoseconds in length, we found that under these constraints K_α can still take a wide range of values (results not shown), and we simply picked a large value within this range.

In the simulations, the solute and the solvent have been separately coupled to temperature baths at 300 K using weak coupling³⁴ with a relaxation time of 0.1 ps. A time step of 2 fs to integrate the equations of motion has been used with all bond lengths and the geometry of the water molecules constrained by SHAKE.³⁵ The twin-range cutoff radii of 0.8 and 1.4 nm with reaction field corrections³⁶ have been applied to treat the nonbonded interactions, interactions within 0.8 nm evaluated every time step through a charge group-based pair-list updated every five steps, the intermediate range interactions within 1.4 nm evaluated at pair-list updates, and the long-range electrostatic interactions beyond 1.4 nm approximated by a dielectric reaction field with a dielectric permittivity of 61. In the sampling simulations, data have been recorded every 0.5 ps for analysis.

The GB1 Peptide. This system consists of the GB1 peptide²⁸ solvated in a box of explicit water. The starting structures for GB1 have been taken from residues 41–56 in the structure of protein G³⁷ (protein data bank or PDB ID 1GB1). This solute has been solvated in a cubic box containing pre-equilibrated SPC³³ water molecules. The box side length is 4.39133 nm, allowing any solute atoms to be at least 0.8 nm from the box sides. The box size has been fixed in further constant volume calculations. Water molecules with oxygen closer than 0.23 nm to any non-hydrogen peptide atoms were removed. Then, three of the remaining water molecules were replaced by sodium ions to neutralize the net charge of the peptide. Finally, 2635 water molecules remained. The GROMOS43A1³² force field has been used. The system was initially energy minimized and then

Table 2. Calculated Free Energies (in kJ/mol) of Different Conformational States Relative to the β and PII Conformation^a for the Alanine Dipeptide

	m^b	T_h^b	γ^b	$\Delta G_{\alpha R}$	$\Delta G_{\alpha L}$	ΔG_{C7ax}
I ^c	10	800 (589.9)	1	6.6 ± 0.5	9.4 ± 0.5	16.3 ± 0.8
		1600 (1063.9)	1	6.4 ± 0.5	10.0 ± 0.4	16.9 ± 1.7
	50	800 (726.2)	1	4.8 ± 1.3	7.9 ± 0.5	15.5 ± 2.2
		1600 (745.0)	0.1	7.3 ± 0.8	9.7 ± 0.8	16.6 ± 0.9
		1600 (1253.2)	0.5	5.2 ± 0.4	9.1 ± 0.3	16.1 ± 0.9
		1600 (1417.6)	1	6.2 ± 0.9	11.2 ± 0.4	19.8 ± 2.9
	200	800 (772.8)	1	5.6 ± 3.8	6.4 ± 0.4	11.4 ± 2.1
		1600 (1119.8)	0.1	6.0 ± 0.6	8.7 ± 1.4	14.5 ± 3.8
		1600 (1468.5)	0.5	6.0 ± 1.1	10.1 ± 0.9	16.4 ± 3.3
		1600 (1536.4)	1	4.0 ± 2.2	7.4 ± 4.8	13.7 ± 2.0
1000	1000	1600 (1392.3)	0.1	5.4 ± 2.3	8.1 ± 3.3	16.2 ± 4.2
		1600 (1546.6)	0.5	6.6 ± 2.5	13.3 ± 6.1	16.9 ± 3.7
		1600 (1576.7)	1	4.6 ± 2.5	9.2 ± 0.8	13.3 ± 3.7
II ^c	× 50	800 (756.9)		12.1 ± 2.3	14.9 ± 0.1	25.2 ± 1.7
		1600 (1508.1)		11.9 ± 1.5	22.9 ± 0.4	35.5 ± 2.2
III ^c		300		5.3 ± 1.8	9.9 ± 1.4	15.2 ± 3.0

^aThe four different conformations are defined on the basis of the (φ, ψ) torsional angles: the β and PII conformation ($-180^\circ < \varphi < 0^\circ$ and $25^\circ < \psi < 180^\circ$), the αR conformation ($-180^\circ < \varphi < 0^\circ$ and $-120^\circ < \psi < 25^\circ$), the αL conformation ($0^\circ < \varphi < 180^\circ$ and $0^\circ < \psi < 180^\circ$), and the C7ax conformation ($0^\circ < \varphi < 180^\circ$ and $-170^\circ < \psi < 0^\circ$). ^bIn the temperature-accelerated simulations, m in atom mass units is the mass associated with the virtual variables, T_h in K is the temperature of the stochastic bath for the virtual subsystem and values in parentheses are the actual temperatures of the virtual subsystem, and γ in ps^{-1} is the friction coefficient for stochastic dynamics simulation. ^cResults in section I have been obtained with adiabatic reweighting using the actual temperature of the virtual coordinates as T_h to obtain reweighting. Results in section II have been obtained by increasing the masses of all solute atoms by 50 fold and coupling the solute to a high temperature bath. Results in section III have been obtained from a normal simulation at 300 K.

subjected to five successive 50 ps constant temperature simulations at 60, 120, 180, 240, and 300 K, respectively. In these simulations, the positions of non-hydrogen solute atoms have been restrained to their starting positions by harmonic restraining potentials with force constants of 5000 kJ/mol/nm². Then, four 50 ps simulations with the restraining force constants reduced to 2500, 1000, 500, and 0, all in kJ/mol/nm², were carried out successively. Finally, a further 5 ns equilibrium simulation at 300 K has been performed.

Starting from the resulting “pre-equilibrated” GB1 peptide configuration, we carried out a 70 ns normal MD sampling simulation and a 70 ns accelerated sampling simulation using the ACM-TAMD method. We consider the GB1 system to test ACM-TAMD because the peptide has a stably folded structure²⁸ (a β -hairpin); thus the anisotropic elastic network model (ANM) can be applied to it despite its small size. We used the “pre-equilibrated” starting configuration as the reference structure in ANM to derive the 10 slowest collective modes, with $r_{\text{cutoff}}^{\text{ANM}}$ in eq 10 set to 0.75 nm. In both the normal and the ACM-TAMD simulations, the time evolution of the collective coordinates along these modes has been followed. The time evolution in the normal simulation allows us to verify whether ANM is a suitable approach to obtain collective coordinates for atomic simulations. In the ACM-TAMD simulation, the first two ANM collective modes have been targeted for accelerated sampling. The masses assigned to the respective virtual variables are 4000 in atom mass units. These variables have been sampled using stochastic dynamics simulations with a bath temperature of 1600 K and a friction coefficient of 5.0 ps^{-1} . The force constant K_a coupling the collective and the virtual variables is 10 000 kJ/mol/nm². When choosing these parameters, the same considerations as in the alanine-dipeptide case have been taken, and short simulations of a few picoseconds have been used to verify that the selected values are within working ranges.

Other simulation parameters used for the GB1 system have been the same as those used for the alanine dipeptide system. The simulations have been performed using the GROMOS05 package³² with home-implemented C++ modules to support TAMD and ACM-TAMD.

RESULTS AND DISCUSSIONS

1. Accuracy of Adiabatic Reweighting for the *n*-Butane System. The average temperatures of the physical and the virtual subsystems in the TAMD simulations of *n*-butane are summarized in Table 1. The deviations of T_{virtual} and T_{physical} from the respective bath temperatures of 1600 and 300 K indicate imperfect adiabatic decoupling between the physical and the virtual subsystems. As has been explained, this must be true because of the coupling interactions in the total potential energy of the extended system (expression 1). However, because the physical subsystem contains more degrees of freedom than the virtual subsystem and also because of the continuous dissipation of heat from the physical subsystem into the stochastic bath, T_{physical} does not deviate substantially from the bath temperature of 300 K, especially for the case with a larger virtual mass (and thus adiabatic decoupling becomes a better approximation). The average coupling interaction energies between the physical and virtual subsystems are also given in Table 1. They are far more close to $(1/2)K_B T_{\text{physical}}$ than to $(1/2)K_B T_{\text{virtual}}$. This and the closeness between T_{physical} and the bath temperature of 300 K suggest that eqs 12–14 that estimate canonical distributions and averages of the physical subsystem through TAMD can be good approximations despite the lack of absolute or exact adiabaticity.

To test that, the distributions of the torsional angle around the central C–C bond and the interatomic distances between the first and the third atoms and between the first and the fourth atoms have been estimated from the TAMD simulations

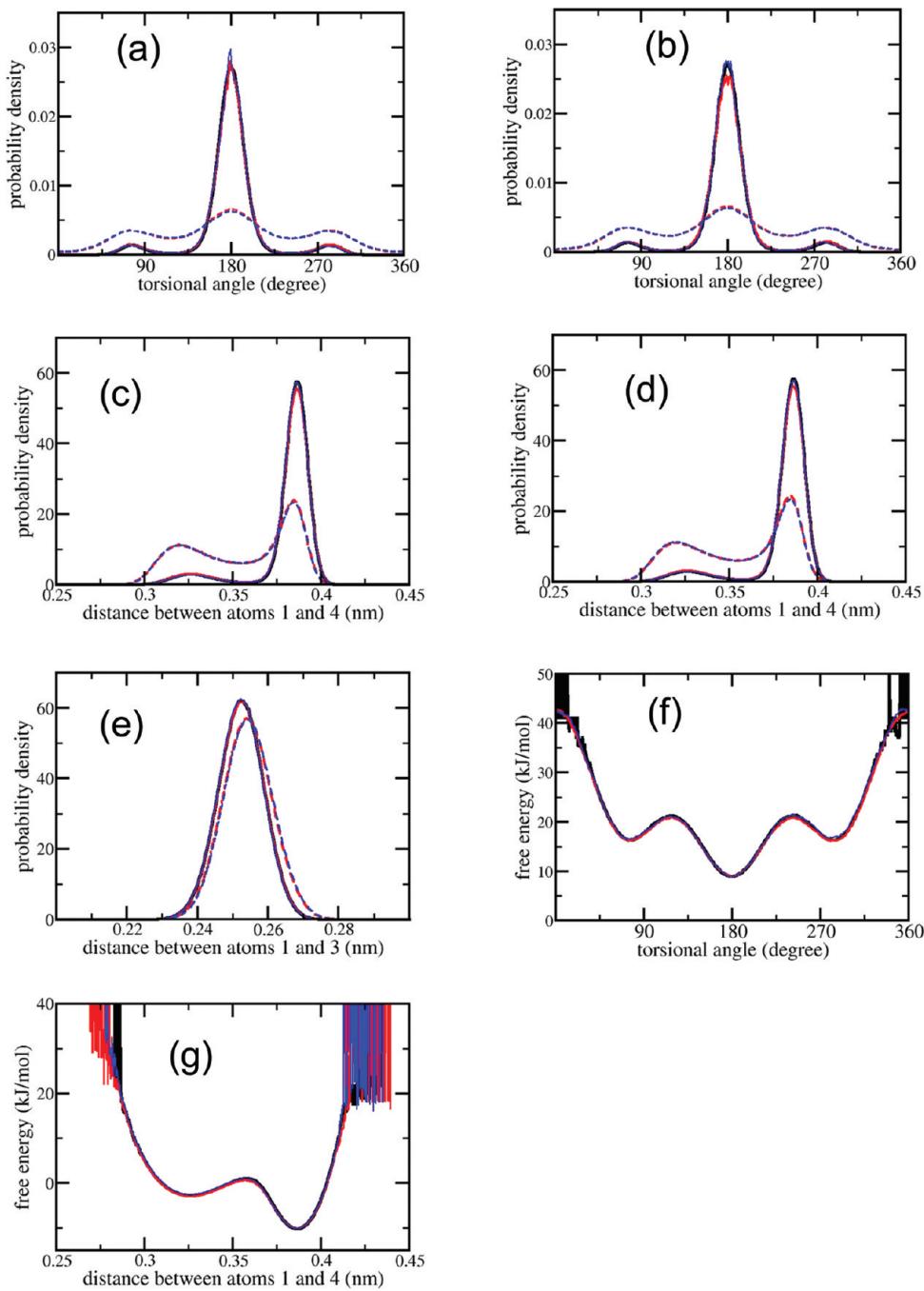


Figure 1. Distributions of geometric parameters (a–e) and free energy surfaces (f,g) of *n*-butane from the normal and the temperature-accelerated simulations. Solid black lines, the normal simulation; red lines, the temperature-accelerated simulation with the virtual mass $m = 1$; blue lines, the temperature-accelerated simulation with the virtual mass $m = 5$. For the accelerated simulations, dashed lines represent the raw distributions, and solid lines represent reweighted canonical results based on adiabatic decoupling. The solid lines, especially the black and the blue lines, mostly fall on top of each other. b and d show results obtained with $\gamma_{\text{physical}} = 2.5 \text{ ps}^{-1}$. The other subfigures show results with $\gamma_{\text{physical}} = 1.5 \text{ ps}^{-1}$.

using eqs 12–14 and compared with the corresponding canonical distributions obtained from the normal SD simulation. The results are shown in Figure 1. For these distributions, the TAMD simulations of the same virtual mass but with different γ_{physical} values yield essentially the same results, so mostly only the results with $\gamma_{\text{physical}} = 1.5 \text{ ps}^{-1}$ are shown (Figure 1a, c, and e). As expected, the unweighted raw distributions from TAMD are significantly different from the canonical ones. This is the case for the distributions of the torsional angle (Figure 1a) and of the distance between atoms 1

and 4 (Figure 1c), which is or directly depends on the coordinate targeted for accelerated sampling. It is also the case for the distance between atoms 1 and 3 (Figure 1e), which does not directly depend on the accelerated coordinate. On the other hand, the distributions obtained through adiabatic reweighting almost exactly reproduce the respective canonical distributions. This is especially true for the TAMD simulations with the virtual mass $m = 5$. Even with $m = 1$, the adiabatic approximation still closely reproduces the exact results, only that the highest peaks in the torsional angle distribution and 1–

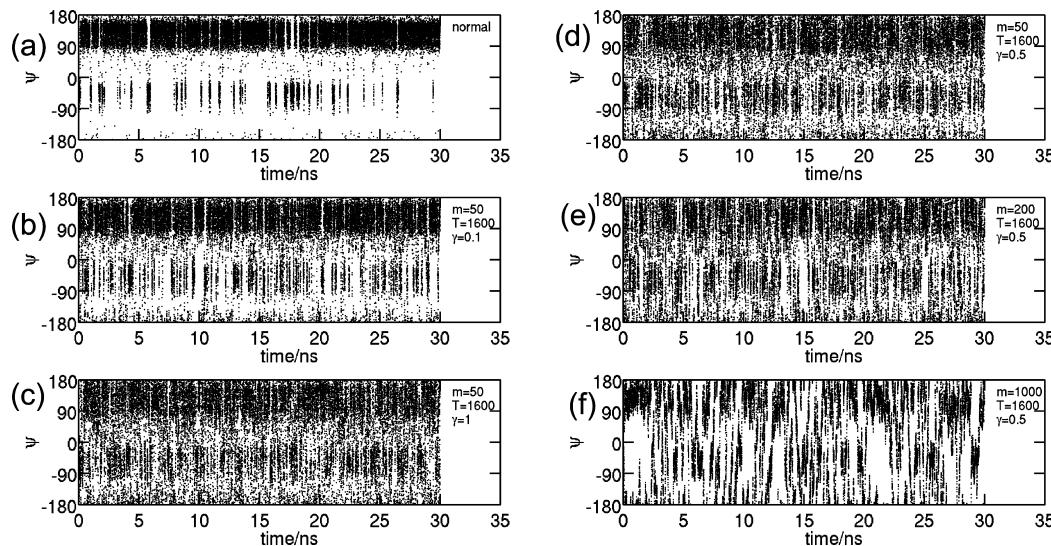


Figure 2. Time trajectories of the dihedral angle ψ in the alanine dipeptide simulations. (a) The normal simulation at 300 K. (b–f) TAMD simulations with different masses (m , in atom mass unit), temperatures (T , in K), and friction coefficients (γ , in ps^{-1}) for the virtual variables as labeled in respective plots.

4 distance distribution are slightly underestimated (to show that these small deficiencies were reproduced with different γ_{physical} , the results with $\gamma_{\text{physical}} = 2.5 \text{ ps}^{-1}$ are also shown for these distributions in Figure 1b and d, respectively). This is probably caused by the small increases in the temperature T_{physical} with $m = 1$. Figure 1f and g show free energy surfaces calculated from the distributions shown in Figure 1a and c, respectively. The TAMD results closely reproduced the canonical free energy surfaces. For $m = 5$, the agreements can be considered as exact. For $m = 1$, small deviations can be noticed and may be explained by the slightly higher T_{physical} in the TAMD simulations.

Although the *n*-butane system is used here not to demonstrate accelerated sampling, it is still worth noting that, through the TAMD simulations, the high free energy region near the *cis* transition barrier on the torsional free energy surface can be determined accurately (statistical uncertainty less than 0.5 kJ/mol from a 50 ns simulation). This is not possible with the 50 ns normal SD simulation, and to obtain the height of the *cis* transition barrier using normal SD would require several orders of magnitude longer simulation time.

2. Acceleration of Sampling for Alanine Dipeptide. In Figure 2, the time evolution of the torsional angle ψ in normal and TAMD simulations with different parameters are compared. Figure 2a shows that in the normal simulation, the transitions between different minima on the Ramachandran (φ, ψ) plane are indeed rare events on the nanosecond time scale. The transitions are mainly between the β and PII region with positive ψ and the α -region with negative ψ . Figure 2b,d show results of TAMD simulations with the virtual variables associated with a mass of 50 and simulated at 1600 K by stochastic dynamics. The transitions are indeed significantly accelerated by TAMD.

Results in Figure 2b–d have been obtained by using different friction coefficients γ on the virtual variables. There are two aspects for the effects of γ . The first aspect is that γ represents the strength of coupling of the virtual variables with the external temperature bath for stochastic dynamics. Because of imperfect adiabaticity, energy flows from the higher temperature virtual subsystem into the lower temperature physical subsystem,

causing the actual averaged temperature of the virtual system to be always below the bath temperature of 1600 K. The larger the γ , the stronger the coupling with the external bath and the higher (i.e., closer to bath temperature) the actual temperature. The higher temperature may result in more significant acceleration effects on sampling. The second aspect for the effects of γ is that γ influences the diffusion rates of the virtual variables (and thus the collective variables of the physical system). For the simulations shown in Figure 2b–d, the actual averaged temperatures are 745, 1253, and 1418 K for $\gamma = 0.1$, 0.5, and 1 ps^{-1} , respectively, in agreement with the expectation that a larger γ yields averaged temperatures closer to the bath temperature. The higher actual temperatures for $\gamma = 0.5 \text{ ps}^{-1}$ (Figure 2d) and $\gamma = 1 \text{ ps}^{-1}$ (Figure 2c) lead to more significant acceleration effects on sampling as compared with $\gamma = 0.1 \text{ ps}^{-1}$ (Figure 2b) despite the slower diffusions associated with the larger γ cases. As for the comparisons between $\gamma = 0.5 \text{ ps}^{-1}$ and $\gamma = 1 \text{ ps}^{-1}$, the different diffusion rates seem to overcompensate the effects of the different temperatures, leading to faster sampling with $\gamma = 0.5 \text{ ps}^{-1}$.

Figure 2d–f are presented to show the effects of different masses taken by the virtual variables. Larger masses make adiabatic decoupling a better approximation. See Table 2, which summarizes the results of the alanine dipeptide simulations. It is shown there (section I, the third column) that when the extended variables take larger masses, the actual temperatures associated with them are closer to the bath temperature, indicating slower energy exchanges between the virtual and the physical subsystems. On the other hand, larger masses lead to slower diffusions of the virtual variables. Thus, going from $m = 50$ to $m = 200$ and $m = 1000$, the acceleration effects on sampling become less significant given the same bath temperature and γ (Figure 2d–f).

Again, we note that nonadiabatic effects are quite obvious from the γ - and mass-dependences of the averaged temperatures of the virtual subsystem and should be taken into account to interpret the varied acceleration effects on sampling. Such a noticeable nonadiabatic effect on the temperature of the virtual subsystem can, however, be mainly attributed to the virtual subsystem containing only two degrees of freedom.

Because of the large number of degrees of freedom of the physical system, the effects of nonadiabaticity on the temperature of the physical subsystem are too small to be observed: we found no statistically significant differences between TAMD and normal MD in the averaged temperatures or the temperature scaling factors for weak coupling of the physical system. This is true whether we only look at the solute or look at both the solute and the solvent. The small nonadiabatic effects on the physical subsystem will be further discussed later by considering ensemble distributions and averages of various interactions.

Taking the above comparisons together, the sampling in the (ϕ, ψ) plane can be easily accelerated by more than an order of magnitude using TAMD, as indicated by comparisons between Figure 2c–e and a. When choosing parameters for TAMD, some general considerations need to be taken into account, but fine-tuning of the parameters is not required.

3. Conformational Free Energy Differences for the Alanine Dipeptide.

Figure 3a–c shows the (ϕ, ψ) free energy

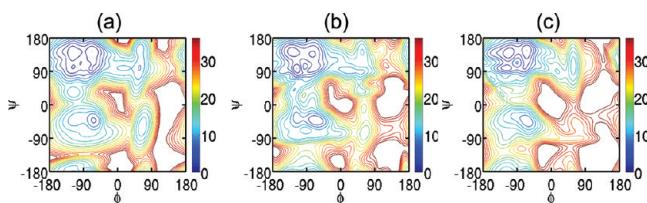


Figure 3. Free energy surfaces along the (ϕ, ψ) coordinates for alanine dipeptide. They have been derived from Gaussian mixture probability densities as described in the main text using (a) the normal simulations; (b) the TAMD simulation with $m = 50$ atom mass units, $T_h = 1600$ K, and $\gamma = 0.5$ ps $^{-1}$ for the virtual variables; and (c) the same as b but $\gamma = 1$ ps $^{-1}$. Adiabatic reweighting has been considered for b and c. The interval between neighboring contour lines is 2 kJ/mol. The maximum contoured free energy level is 36 kJ/mol.

surfaces constructed from the normal simulation and two TAMD simulations (using eqs 8, 15, and 16 assuming adiabatic decoupling and using the actual averaged temperature of the virtual variable as T_h). The probability density functions used to derive the free energy surfaces have been estimated as the sum of 25 Gian functions by the maximum likelihood method²⁷ (the 2π periodicity of the torsional angles has been considered in this case). The free energy surfaces in general agree well with each other. Compared to the TAMD simulations, the surface reconstructed using the normal simulation is expected to be more accurate around local minima because of concentrated sampling in these regions but less accurate in higher energy regions (including barrier regions) because of increased statistical uncertainties caused by a smaller number of sampled configurations in these regions.

To obtain more quantitative comparisons, free energy differences between different conformational states defined using (ϕ, ψ) have been computed from the normal and the TAMD simulations through counting the number of sampled configurations in the different (ϕ, ψ) regions. The results are presented in Figure 4 and Table 2. The partitioning of the (ϕ, ψ) plane into different conformational states is given in the footnotes of Table 2. For the TAMD simulations, the sampled configurations have been adiabatically reweighted according to eqs 13 and 14 when carrying out counting. The free energy surfaces needed for reweighting have been determined using

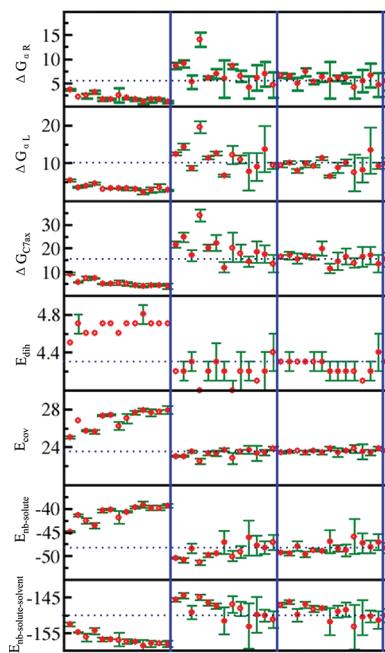


Figure 4. From top to bottom, calculated free energies (in kJ/mol) of different conformational states of the alanine dipeptide relative to the β and PII conformation and ensemble averages of various interaction energies (in kJ/mol) from the alanine dipeptide simulations. See also, footnotes of Tables 2 and 3. In each subfigure and from left to right, three groups of averages are separated by the solid blue vertical lines, each group containing results from 13 TAMD simulations with different combinations of parameters, as specified in Tables 2 and 3. The first group contains unweighted averages. The second group contains reweighted averages with the temperature of the stochastic bath used as T_h . The third group contains reweighted averages with the actual average temperature of the virtual subsystem used as T_h . In each panel, the right-most point and the blue dotted line correspond to the result from the normal 300 K simulation. Error bars have been obtained as standard variations between 3×10 ns blocks of simulation.

maximum likelihood density estimation. Twenty-five Gaussian functions have been used for density estimation, but the results are insensitive to reasonable changes in this number. To estimate statistical uncertainties, each simulation has been partitioned into 3×10 ns blocks. The corresponding uncertainties have been estimated as standard deviations between results of the three blocks.

In Figure 4, the relative free energies computed with and without adiabatic reweighting of the sampled configurations are compared with the results from the normal simulation. From Figure 4, one can see that without adiabatic reweighting, the free energy differences are systematically overestimated. This confirms that the conformational distribution sampled by TAMD deviates nontrivially from the canonical Boltzmann distribution.

To obtain the adiabatic reweighting results, two choices have been considered. The first is to derive the reweighting factors by taking the temperature of the stochastic bath for the virtual variables as T_h in eqs 8, 13, and 14. The second choice is to use the actual temperature of the virtual variables as T_h . The first choice led to results that are significantly improved over the unweighted results but still contain systematic errors (Figure 4). These errors are especially prominent when the deviations between the bath and actual temperatures are large due to small

Table 3. Ensemble Averages of Various Interaction Energies^a (in kJ/mol) from the Alanine Dipeptide Simulations

	m^b	T_h^b	γ^b	E_{dih}	E_{cov}	$E_{\text{nb-solute}}$	$E_{\text{nb-solute-solvent}}$
I ^b	10	800 (589.9)	1	4.3 ± 0.0	23.4 ± 0.1	-49.3 ± 0.2	-147.2 ± 0.6
		1600 (1063.9)	1	4.3 ± 0.1	23.5 ± 0.1	-49.5 ± 0.4	-146.4 ± 0.6
	50	800 (726.2)	1	4.3 ± 0.0	23.6 ± 0.0	-48.0 ± 1.0	-150.0 ± 1.8
		1600 (745.0)	0.1	4.3 ± 0.0	23.4 ± 0.2	-49.7 ± 0.6	-147.1 ± 0.9
		1600 (1253.2)	0.5	4.3 ± 0.1	23.6 ± 0.1	-48.8 ± 0.4	-148.3 ± 0.8
		1600 (1417.6)	1	4.3 ± 0.1	23.4 ± 0.2	-49.0 ± 0.2	-148.1 ± 0.5
	200	800 (772.8)	1	4.2 ± 0.1	23.8 ± 0.2	-47.0 ± 2.4	-151.9 ± 3.8
		1600 (1119.8)	0.1	4.2 ± 0.1	23.4 ± 0.5	-48.5 ± 0.7	-149.1 ± 1.7
		1600 (1468.5)	0.5	4.2 ± 0.1	23.6 ± 0.1	-48.9 ± 1.3	-148.6 ± 2.0
		1600 (1536.4)	1	4.2 ± 0.1	23.8 ± 0.4	-46.0 ± 3.7	-153.4 ± 6.3
1000	1000	1600 (1392.3)	0.1	4.1 ± 0.0	23.5 ± 0.8	-47.4 ± 2.1	-150.7 ± 4.9
		1600 (1546.6)	0.5	4.2 ± 0.1	23.4 ± 0.3	-48.1 ± 1.4	-150.4 ± 1.6
		1600 (1576.7)	1	4.4 ± 0.2	23.8 ± 0.2	-47.1 ± 1.6	-151.4 ± 2.4
II ^c	50 fold	800 (756.9)		8.7 ± 0.1	56.8 ± 0.8	-48.4 ± 2.3	-135.3 ± 4.1
		1600 (1508.1)		17.2 ± 0.5	106.2 ± 1.4	-42.2 ± 1.5	-129.6 ± 1.3
III ^c		300		4.3 ± 0.0	23.6 ± 0.0	-48.3 ± 1.1	-150.2 ± 1.8

^a E_{dih} is the total dihedral angle interaction energy. E_{cov} is the total covalent interaction energy. $E_{\text{nb-solute}}$ is the total nonbonded interaction energy of solute. $E_{\text{nb-solute-solvent}}$ is the total solute–solvent interaction energy. ^bSee corresponding footnote of Table 2. ^cSee corresponding footnote of Table 2.

virtual masses that enhance the nonadiabatic effects and/or small friction coefficients that weaken the coupling with the stochastic temperature bath. The second choice, namely, deriving the adiabatic reweighting factors using the actual temperature of the virtual coordinates as T_h , led to results that mostly agree with the normal simulation results within statistical errors (Figure 4 and Table 2). Thus, the results in Table 2 support the theoretical approximations leading to the adiabatic reweighting formulas and the practical usefulness of these formulas.

In Table 2, we also presented results from an adiabatic decoupling simulation without using collective/virtual variables. In this simulation, all solute masses have been up-scaled by 50 fold and coupled to a higher temperature bath. Accelerated sampling in the (φ , ψ) plane has been observed as expected (results not shown). However, because of the high dimensionality of the overall higher temperature subspace, a free energy surface cannot be constructed, and adiabatic reweighting cannot be applied. Nevertheless, we presented in Table 2 the results obtained by assuming that only the (φ , ψ) coordinates are distorted by the higher temperature, and all other orthogonal degrees of freedom follow the normal canonical distribution. Results in Table 2 suggest that this is a very poor approximation. As has also been suggested in ref 8, except for special cases, adiabatic decoupling simulations considering atomic degrees of freedom to form the higher temperature subspace may be used to explore the conformational space but involve significant biases in their sampled conformational distributions, and such biases are practically difficult to correct.

4. Ensemble Averages for Alanine Dipeptide. The free energy differences in Figure 4 and Table 2 can be considered as ensemble averages of properties that depend only on the accelerated collective variables. In Figure 4 and in Table 3, we also presented ensemble averages of the various interactions that depend on the unaccelerated, orthogonal coordinates as well. The interactions considered include the total dihedral angle energies, the total covalent interaction energies, the total nonbonded interaction energies of solute, and the total nonbonded interaction energies between solute and solvent. Direct averages over the TAMD simulations (the left-most group of results in Figure 4) deviate largely from direct averages

over the normal simulations. For $T_h = 800$ K, the deviations of the above interaction energies in the mentioned order are between 0.2–0.3 kJ/mol, 1.4–2.6 kJ/mol, 3.8–6.8 kJ/mol, and -6.8 to -2.5 kJ/mol, respectively. For $T_h = 1600$ K, the deviations of the same interaction energies in the mentioned order are 0.3–0.5 kJ/mol, 2.0–4.3 kJ/mol, 4.7–9.1 kJ/mol, and -8.4 to -4.7 kJ/mol, respectively. Thus, the first three interactions that involve only solute coordinates are systematically higher in TAMD than in the normal simulation, consistent with the higher temperature of the solute collective coordinates in TAMD. Indeed, Figure 4 shows that the higher the T_h , the larger these deviations. On the other hand, the total solute–solvent interaction is systematically lower in TAMD than in the normal simulation, probably because the solute becomes more flexible in TAMD and easier to accommodate configurations that have more favorable solute–solvent (but less favorable solute–solute) interactions.

Adiabatic reweighting significantly lowers the deviations between the ensemble averages from the various TAMD simulations and from the normal simulation. The resulting deviations are mostly below the respective statistical uncertainties (Figure 4 and Table 3). In addition, the systematically positive or negative deviations that present in the unweighted averages have diminished. This is especially true when the actual temperatures of the virtual variables have been used as T_h in adiabatic reweighting. The deviations of respective interaction energies are -0.3 to 0.1 kJ/mol, -0.6 to 0.2 kJ/mol, -3.1 to 1.9 kJ/mol, and -2.0 to 5.6 kJ/mol if the bath temperature has been used to derive adiabatic reweighting, while the same deviations are -0.2 to 0.1 kJ/mol, -0.2 to 0.2 kJ/mol, -1.4 to 1.3 kJ/mol, and -3.0 to 3.8 kJ/mol if the actual temperature of the virtual variables has been used to derive adiabatic reweighting.

On the basis of results in Figure 2 and Table 2, the parameters $m = 50$ and $\gamma = 0.5 \text{ ps}^{-1}$ can be considered as mass and friction parameters that properly balance between adiabaticity and sampling efficiency. The adiabatically reweighted averages over the TAMD simulation with these parameters yields almost the same averages as the normal simulations (the respective deviations of the various energy terms are 0.0, 0.0, -0.5, and 1.9 kJ/mol, which can be

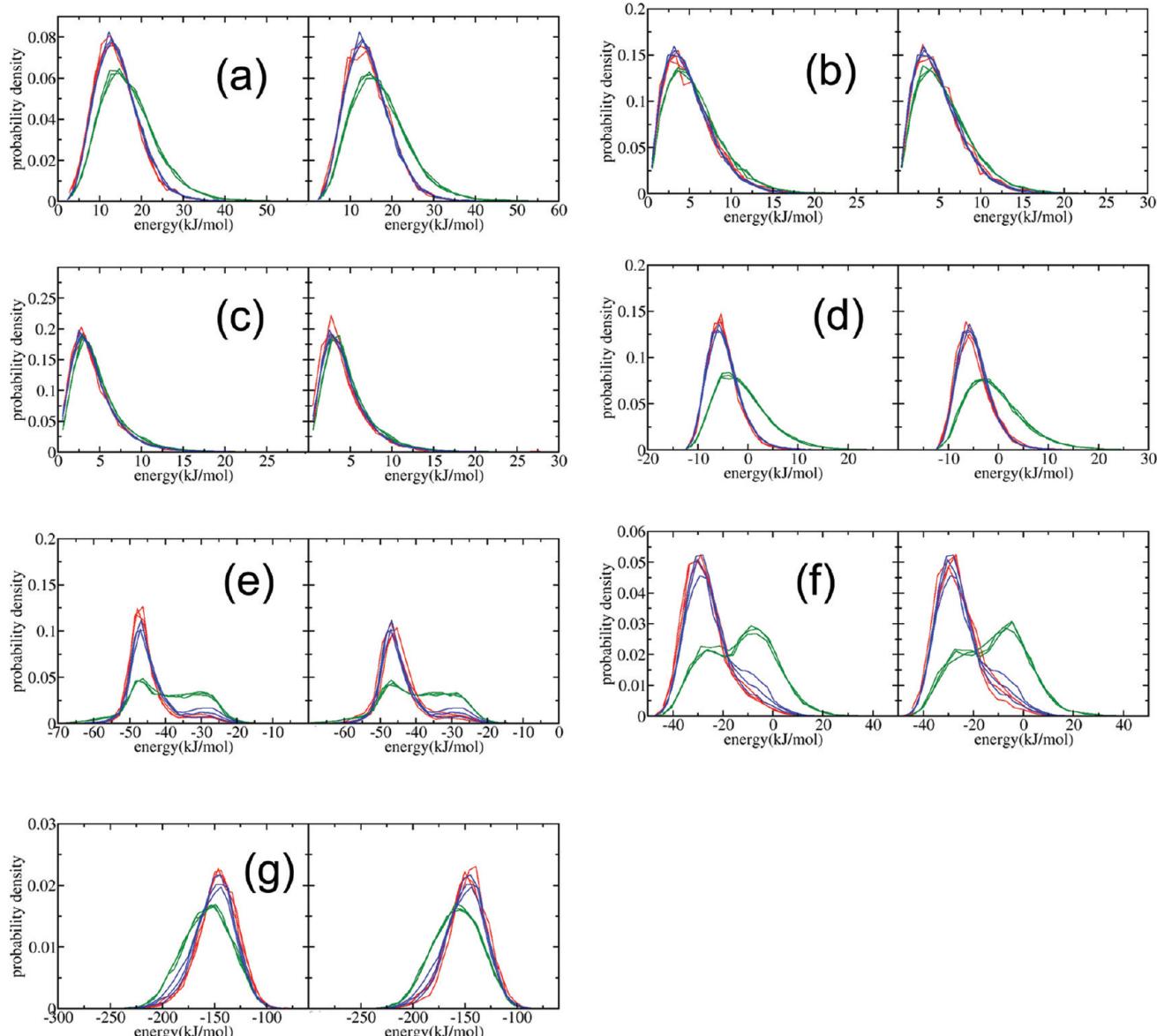


Figure 5. Distributions of various interaction energies of the alanine dipeptide in water. (a) Bond angle energy, (b) improper dihedral angle energy, (c) torsional angle energy, (d) solute van der Waals energy, (e) solute electrostatic energy, (f) total potential energy of solute, and (g) total solute–solvent interaction energy. Blue lines, normal 300 K simulations; green lines, raw distributions from TAMD simulations; red lines, reweighted distributions from TAMD simulations. Left panels: TAMD parameters $m = 50$, $\gamma = 0.5 \text{ ps}^{-1}$, and $T = 1600 \text{ K}$. Right panels: TAMD parameters $m = 200$, $\gamma = 0.5 \text{ ps}^{-1}$, and $T = 1600 \text{ K}$. To present statistical uncertainties, each line in a cluster of three lines represents results of a 10 ns block in a 30 ns simulation.

compared with the deviations of 0.4 kJ/mol, 3.7 kJ/mol, 7.8 kJ/mol, and -6.7 kJ/mol for unweighted averages over the same simulation). In addition, the adiabatic reweighting results produced smaller standard deviations between the three 10 ns simulation blocks than the results from the normal simulation. This again indicates that the reweighting formulation assuming adiabatic decoupling is practically applicable, i.e., without introducing too large statistical uncertainties.

To further assess the accuracy of adiabatic reweighting, the distributions of various interaction energies obtained from the normal simulations and from the TAMD simulations are compared. Figure 5 shows the results for the two sets of TAMD simulations with $m = 50$, $\gamma = 0.5$, and $T = 1600 \text{ K}$ and with $m = 200$, $\gamma = 0.5$, and $T = 1600 \text{ K}$. For most of the interaction

energies, the raw distributions from the TAMD simulations deviate significantly from the distributions yielded by the normal MD simulation. Adiabatic reweighting changes the TAMD distributions nontrivially; the resulting distributions agree with the normal distributions within respective statistical uncertainties. Thus, despite imperfect adiabaticity, adiabatic reweighting can quite accurately reproduce the usual canonical distributions and averages. Again, the first reason is that the total amount of heat transferred into the physical system is very small considering the large number of degrees of freedom of the physical system. The second and perhaps more important reason is that the virtual masses have been chosen to be large enough so that there is enough time for the transferred heat to dissipate into the vast number of physical degrees of freedom. As a result, the average per-degree-of-freedom coupling

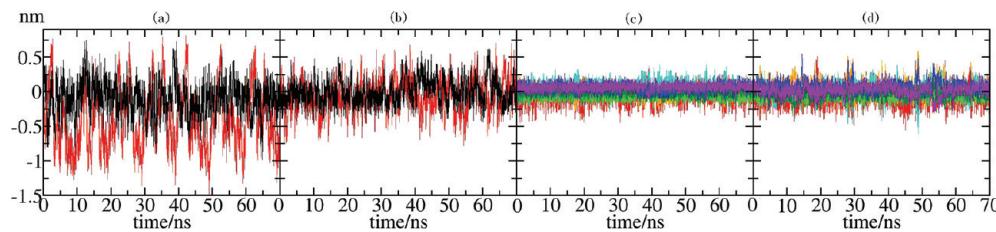


Figure 6. For the GB1 peptide, time evolution of displacements along the slowest collective modes derived from the elastic network model. (a) The first mode. The black line corresponds to the normal simulation and the red line, the ACM-TAMD simulation. (b) The same as a but for the second mode. (c) The third to the ninth modes in different colors from the normal simulation. (d) The same as c but from the ACM-TAMD simulation.

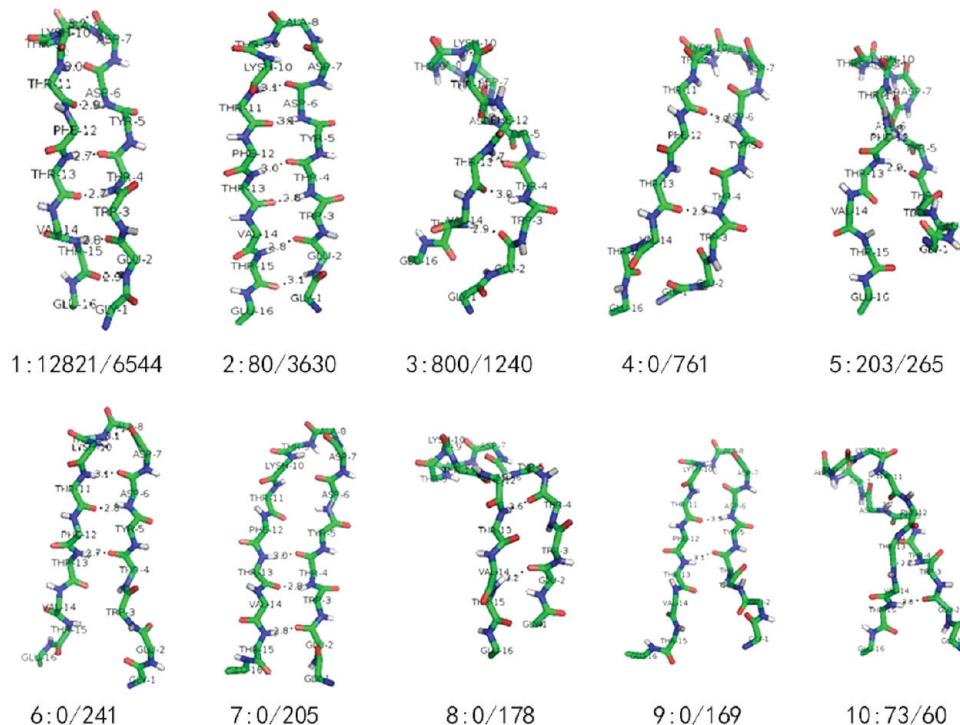


Figure 7. Conformational clusters of the GB1 peptide from the normal simulation and the ACM-TAMD simulation. Clustering is based on positional RMSD of all backbone atoms with a cluster radius of 0.15 nm. The representative conformer for each of the largest 10 clusters is shown. Three numbers are given below each conformer: the first, the rank of the cluster; the second, the number of conformers from the normal simulation in this cluster; and the third, the number of conformers from the ACM-TAMD simulation in this cluster.

interaction energies between the virtual and the physical variables are 1.26 and 1.24 kJ/mol, respectively, for the two sets of TAMD parameters, which vary close to the ideal value of $(1/2)K_B T_r$.

5. Accelerated Sampling of the G1 Peptide. Figure 6 shows the time-dependent fluctuations of the lowest nine ANM modes in the normal and the ACM-TAMD simulations. The eigenvalues of the lowest two modes are 0.01064 and 0.04376, respectively. The eigenvalues of the third to the ninth modes are 0.08481, 0.15751, 0.24768, 0.33165, 0.38140, 0.52146, and 0.55715, respectively. From the curves representing the normal simulation that is independent of ANM, we can see that fluctuations along the first ANM mode are indeed the largest in the atomic simulation (Figure 6a). Fluctuations along the second ANM mode are the second largest (Figure 6b) as well. As one goes to higher order modes, the fluctuations get smaller, as shown for the third to the ninth modes in Figure 6c. In addition, we have observed that the fluctuations along the slowest ANM modes were largely uncorrelated (results not shown). This justifies the idea of ACM, namely, targeting the

first few coarse-grained ANM collective modes for accelerated sampling around well-folded structures.

Comparing the ACM-TAMD curves with the curves representing normal simulations in Figure 6 indicates that with $T_h = 1600$ K, sampling along the first ANM mode is significantly amplified in ACM-TAMD (see Figure 6a). Sampling along the second ANM mode is also amplified, but to a lesser extent as compared with the first mode (see Figure 6b). This is consistent with the expectation of stiffer motions along the second mode. Motions along the third to the ninth modes still remain restricted as in the normal simulation, although occasional larger fluctuations in some modes are observed (Figure 6d).

To further analyze the distribution of conformations in the different simulations, conformations sampled every 5 ps by the normal simulation and by ACM-TAMD are pooled together and clustered. Figure 7 shows the largest 10 conformation clusters. For each cluster, the total number of conformers as well as the numbers of conformers from the normal and the ACM-TAMD simulations have been given. The conformations sampled by the normal simulation are highly concentrated: the

normal simulation contributed more than 12 800 conformers to cluster 1 and contributed more than 200 conformers to only three clusters (clusters 1, 3, and 5). The ACM-TAMD simulation sampled different conformational clusters more evenly. The first three clusters all contain more than 1000 conformers from ACM-TAMD, and there are seven clusters containing more than 200 conformers from ACM-TAMD. More importantly, 5 out of the top 10 clusters, including clusters 4 and 6–9, contain conformers only from the ACM-TAMD simulation but not from the normal simulation. Besides that, the second largest cluster contains more than 3600 conformers from ACM-TAMD but only 80 from the normal simulation. Thus ACM-TAMD sampled much more extensively in the conformational space because of the accelerated and amplified motions along the two collective ANM modes.

6. Free Energy Surfaces of the GB1 Peptide. We estimated the free energy surfaces from the 70 ns sampling simulations using the first two collective ANM modes as reaction coordinates. The probability density functions were estimated as mixtures of 10 Gaussian functions with the maximum likelihood and expectation maximization approach.²⁷ Figure 8a

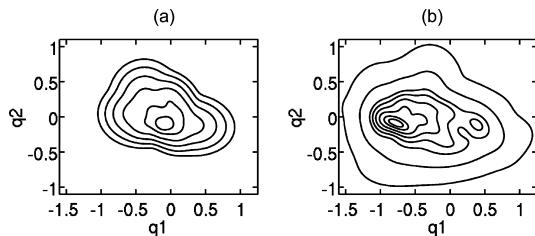


Figure 8. For GB1 peptide, free energy surfaces along the first (q_1) and the second (q_2) slowest ANM collective modes. They have been derived from probability densities estimated as Gaussian mixtures as described in the main text. (a) Normal simulation. Contour lines correspond to levels of 1 kJ/mol, 3 kJ/mol, 7 kJ/mol, 12 kJ/mol, 18 kJ/mol, and 25 kJ/mol, respectively. (b) ACM-TAMD simulation with adiabatic reweighting considered. Contour lines represent levels of 1 kJ/mol, 3 kJ/mol, 7 kJ/mol, 12 kJ/mol, 18 kJ/mol, 25 kJ/mol, 60 kJ/mol, and 155 kJ/mol, respectively. The outmost levels in a and b just encompass all sampled configurations in the respective simulations.

and b show results obtained using the normal simulation and the ACM-TAMD simulation, respectively. As expected, the ACM-TAMD covered a much more extensive region of the free energy surface. On the other hand, the two free surfaces differ significantly. The minima are shifted toward a negative value along the first mode in the ACM-TAMD results. In addition, the ACM-TAMD surface shows a more complex landscape, while the surface constructed from the normal simulation indicates a simple landscape with a single minimum located near the “pre-equilibrated” starting conformation (namely, the origin of the collective coordinates because this starting conformation has been used as the reference to compute the ANM coordinates). This is consistent with the conformation cluster analysis showing that more rich and diverse conformations have been sampled by the ACM-TAMD simulation.

The results in Figure 8b for the TAMD simulation have been obtained using the adiabatic decoupling assumption. As has been discussed before, absolute adiabatic decoupling should not be expected. However, for the same reasons discussed for the *n*-butane system and the alanine-dipeptide system, the effects of imperfect adiabatic decoupling are not expected to be a major

source of error in these results. For example, the TAMD simulation and the normal simulation show no noticeable differences in their average temperatures or temperature scaling factors. The average coupling interactions between the physical and the virtual collective variables are 1.46 and 1.68 kJ/mol for the first and second modes, respectively, being comparable to results for the *n*-butane system shown in Table 1. In addition, each of the two targeted physical collective modes is a combination of 16 atomic coordinates (48 degrees of freedom), which are involved in diverse interactions that can rapidly dissipate the small amount of heat transferred due to imperfect adiabatic decoupling. In fact, if the accuracy of adiabatic reweighting for the *n*-butane system could be generalized to the GB1 system, we would expect that the errors due to imperfect adiabatic decoupling to be negligibly small for results in Figures 8b. Then, the differences between Figure 8a and b may be attributed to improved sampling in TAMD. The following comparisons of conformations sampled by the simulations with experimentally derived structural models of GB1 seem to support this indication.

7. Improved Conformational Distribution Sampled by ACM-TAMD. Because of their limited lengths of 70 ns, both the normal simulation and the ACM-TAMD simulation can only sample locally in GB1 peptide’s conformational space, that is, near the native β hairpin structure (see conformers in Figure 7, this also justifies the application of the ANM model). Thus, one may ask whether the above differences between the conformational distributions sampled by the two simulations are indeed due to the fact that sampling by normal simulation is more severely affected by the limited time scale of the simulation and that ACM-TAMD has improved over it, or the differences are just due to the fact that the two simulations happened to have been sampling different local regions in the conformational space. We try to address this issue through the following analyses.

First, the conformational clustering analysis, which does not depend on the projection of the high dimensional conformational space onto a low dimensional space, indicates that the local regions of conformational space sampled by the normal simulation have been covered by ACM-TAMD sampling. Thus, the remaining question is whether the more extensive sampling by ACM-TAMD is a real improvement over the normal simulation.

To look into this, we have compared the conformers sampled in the simulations with experimentally derived structures of the GB1 peptide in the protein data bank. From 27 PDB entries (the PDB IDs are 1EM7, 1FCC, 1FD6, 1GB1, 1GB4, 1MPE, 1PGA, 1PGB, 1Q10, 1QKZ, 1UWX, 2GB1, 2GI9, 2I2Y, 2I38, 2IGG, 2JU6, 2KLK, 2KN4, 2ON8, 2ONQ, 2QMT, 2RMM, 3FIL, 3GB1, 3MP9, and 3UI3), we retrieved 441 structural models that are well-folded. Each of the representative conformations shown in Figure 7 has been compared with each of the 441 experimentally derived models, the pairwise RMSD of backbone atom positions shown in Figure 9. Most of the experimental models have the smallest RMSD from the representative conformation of cluster 2, the respective RMSD values being below 0.15 nm. A few of the experimental models have the smallest RMSD from the representative conformation of cluster 1. Most members of cluster 2 come from the TAMD simulation. The other conformation clusters resemble the experimental models less than clusters 1 and 2. This is not unexpected because these clusters have much lower probabilities in the simulated ensembles, while the experimental

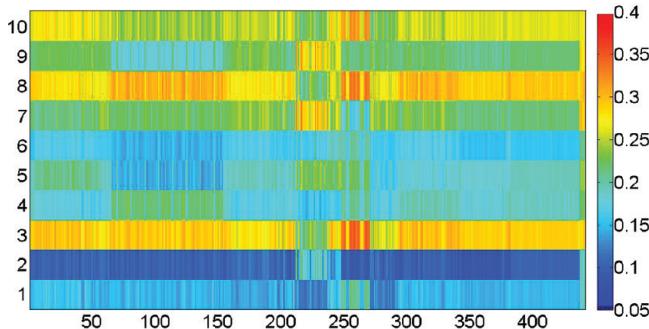


Figure 9. Root mean square deviations (in nm) of backbone atom positions between conformations sampled in the GB1 simulations and 441 experimentally derived structures retrieved from the protein data bank. Each row corresponds to the representative conformer of a conformation cluster, while each column corresponds to an experimental model. The clusters are ranked the same as in Figure 7.

models have been derived on the basis of ensemble-averaged observable properties. Among the remaining clusters that resemble many (clusters 4 and 6) or some (clusters 5, 7, and 9) experimental models, only cluster 5 contains members sampled by the normal simulation. Thus, the set of conformations sampled by TAMD appear to better represent the conformational diversity of GB1 than the set of conformations sampled by the normal simulation of the same length. Interestingly, the two clusters that are most far away from the experimentally derived models (clusters 3 and 8) both represent globally twisted or bent hairpin structures. As results, they are also the most far away in terms of RMSD from clusters 1 and 2, which dominated the sampled conformational ensembles.

To include the entire sets of sampled configurations instead of only the representative conformations of the clusters in the analysis, we calculated the time trajectories and distributions of two types of root-mean-square deviation of atomic positions (RMSD) from the simulations. One type is the RMSD from the “pre-equilibrated” starting structure of the simulation, the results shown in Figure 10a and b. The other type is the RMSD from the experimental structure that has been used to set up the simulation from the beginning, the results shown in Figure 10c and d. The positions of backbone atoms have been considered for RMSD calculations. The RMSD between the “pre-equilibrated” starting structure and the experimental structure themselves is 0.22 nm. For ACM-TAMD, the distributions of the two types of RMSD with and without adiabatic reweighting are both given. As before, adiabatic reweighting, if considered, has been based on eqs 13 and 14 using the reconstructed free energy surface shown in Figure 8b.

For the RMSD from the “pre-equilibrated” starting configuration, the normal simulation yielded a far narrower distribution than ACM-TAMD. In addition, the peak position in the distribution of the normal simulation has a much smaller RMSD value (0.12 nm). The unweighted ACM-TAMD distribution is broader with not a peak but a flat top. After adiabatic reweighting, the distribution is narrowed but still not as narrow as the distribution of the normal simulation. More noteworthy, adiabatic reweighting caused the peak position to shift toward a much larger RMSD value (0.2 nm) from the “pre-equilibrated” starting configuration.

For the RMSD from the experimental structure used for initial set up, the normal simulation produced a distribution peaked at 0.2 nm. Thus, on average, the conformations sampled

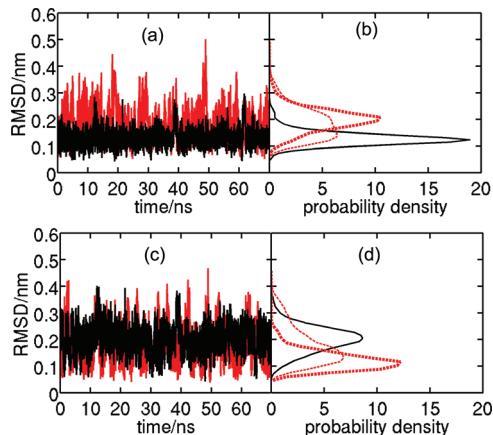


Figure 10. Time trajectories and distributions of the root-mean-square deviations (RMSD) in the normal (black lines) and ACM-TAMD (red lines) simulations. a and b are for the RMSD from the pre-equilibrated starting structure. c and d are for the RMSD from the experimental structure used for set up. Backbone atoms have been considered for RMSD calculations. In b and d, showing the distributions, the thinner dashed lines in red are unweighted ACM-TAMD results, and the thicker dashed lines in red are ACM-TAMD results with adiabatic reweighting.

by the normal simulation are significantly closer to the “pre-equilibrated” starting structure than to the experimental structure used for set up. On the other hand, the ACM-TAMD simulation displayed a reverse trend. Despite the more diverse conformations sampled, the distributions of RMSD from the experimental structure for setup generated by ACM-TAMD peaked at a much smaller RMSD value as compared with the normal simulation. Adiabatic reweighting caused the distribution to be narrower, the peak moving toward an even smaller RMSD value (0.1 nm). In other words, the conformations sampled by the ACM-TAMD are on average significantly closer to the experimental structure used for set up than to the “pre-equilibrated” starting structure.

Considering that both the normal simulation and the ACM-TAMD simulation were started from the same configuration, used the same force field model, and simulated the same solution environment, we may infer that the results of the normal simulation being closer to the “pre-equilibrated” starting structure than to the experimental structure used for set up is not due to force field biases or different environments between simulation and experiment but due to insufficient sampling in the 70 ns normal simulation even for such local conformational fluctuations. In addition, the results in Figures 7, 9, and 10 taken together support the conformational distribution sampled by ACM-TAMD with adiabatic reweighting improving over the distribution sampled by the normal 300 K simulation.

CONCLUSIONS

In accelerated sampling such as TAMD or adiabatic decoupling simulations, which are based on elevated temperatures and targeting a small number of degrees of freedom, equilibrium conformation distributions may exist but do not follow the canonical Boltzmann distribution. Thus, directly averaging over the sampled conformations does not give correct ensemble averages.

It is shown that by making the adiabatic decoupling assumption, the sampled conformations can be reweighted to obtain the correct ensemble averages. The reweighting relies on

constructing the low-dimensional free energy surfaces along the accelerated coordinates. The free energy surface is to be derived from the density probability function, which can be estimated from the sampled configurations using maximum likelihood. More specifically, the density can be estimated using the standard approach of Gaussian mixture densities and expectation maximization to determine the Gaussian parameters and weights. In practice, this approach is possible only for the estimation of low dimensional densities. Thus, the proposed adiabatic reweighting scheme is especially useful for adiabatic TAMD targeting only a few collected coordinates or variables for accelerated sampling.

Previously, we have proposed the amplified collective motion or ACM method. In ACM, collective modes from the coarse-grained elastic network model or ANM are employed to accelerate atomic simulations. Motions in the subspace spanned by the collective modes are coupled to a higher temperature bath. Because the physical atomic velocities are directly perturbed, the original ACM is essentially a nonequilibrium simulation technique. Here, we combine the idea of ACM and the idea of extending the phase space variable in meta-dynamics and TAMD. The slowest ANM modes are targeted for acceleration as in ACM, but the virtual variable approach of TAMD is employed to avoid directly changing the physical velocities. By this ACM-TAMD approach, sampled conformations can be adiabatically reweighted to recover the canonical distribution and to obtain correct thermodynamic averages.

To assess the accuracy of the adiabatic decoupling approximation, we considered *n*-butane in the gas phase, for which we found that conformation distributions obtained from TAMD simulations and using adiabatic reweighting can accurately reproduce the actual canonical distributions.

To test and demonstrate the practical applicability of the proposed approaches, we then used explicitly solvated alanine dipeptide and GB1 peptide as model systems. For the alanine dipeptide, we targeted the peptide backbone Ramachandran dihedral angles for accelerated sampling using TAMD. It is shown that sampling can be accelerated by more than an order of magnitude. We additionally show that in spite of the accelerated sampling, distributions of the accelerated coordinates as well as canonical distributions and averages of various interaction energies can be recovered with satisfactory accuracy under the adiabatic approximation. We also show that for parameter choices, some general considerations to balance between adiabaticity and diffusion along the collective coordinates are necessary. Nevertheless, not much fine-tuning of parameters is required to achieve significant acceleration over normal simulations.

For the GB1 system, it has been verified that a normal atomic simulation indeed produced the largest fluctuations along the slowest ANM modes. An ACM-TAMD simulation targeting the slowest two ANM modes for accelerated sampling has been compared with a normal simulation. Both simulations sample the conformational space locally. The ACM-TAMD simulation enclosed the local regions sampled by the normal simulation, but the reverse is not true. More importantly, comparing RMSD distributions from the simulations suggested that the conformation distribution sampled by ACM-TAMD, after adiabatic reweighting, significantly improves over the distribution sampled by the normal simulation that suffers severely from insufficient sampling.

In summary, when simulation time or the efficiency of sampling is the bottleneck, TAMD and ACM-TAMD with

adiabatic reweighting may demonstrably improve over conventional MD in terms of representing the canonical distribution of molecular conformations, a fundamental goal of molecular simulations. In future work, we plan to further investigate the accuracy of adiabatic reweighting in other important applications such as free energy calculations, which can potentially benefit from the improvement of sampling by TAMD and ACM-TAMD. We also plan to apply the ACM-TAMD approach to larger globular or multidomain proteins, not only for exploring the conformational space but also for improved prediction of conformational distributions and thermodynamic ensemble averages.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hyliu@ustc.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully thank Professor van Gunsteren for providing the GROMOS05 package and for insightful discussions about adiabatic decoupling and enhanced sampling. Financial supports have been provided by the Chinese Natural Science Foundation (grant numbers 30970560 and 21173203).

■ DEDICATION

This work is dedicated to Prof. van Gunsteren on the occasion of his 65th birthday.

■ REFERENCES

- (1) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Ostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.A.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064–4092.
- (2) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulter, K. *Structure* **2009**, *17*, 1295–1306.
- (3) Schlick, T.; Colleparo-Guevara, R.; Halvorsen, L. A.; Jung, S.; Xiao, X. Q. *Rev. Biophys.* **2011**, *44*, 191–228.
- (4) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142–147.
- (5) van Gunsteren, W. F.; Dolenc, J. *Biochem. Soc. Trans.* **2008**, *36*, 11–15.
- (6) Christen, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157–166.
- (7) Bolhuis, P. G.; Dellago, C. *Rev. Comput. Chem.* **2011**, *27*, 111–210.
- (8) Kunz, A.-P. E.; Liu, H.; van Gunsteren, W. F. *J. Chem. Phys.* **2011**, *135*, 104106.
- (9) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893–2906.
- (10) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- (11) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (12) Schlitter, J. M.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.
- (13) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- (14) Di Nola, A.; Roccatano, D.; Berendsen, H. J. C. *Proteins* **1994**, *19*, 174–182.
- (15) Mangoni, M.; Roccatano, D.; Di Nola, A. *Proteins* **1999**, *35*, 153–162.
- (16) Zhang, Z.; Shi, Y.; Liu, H. *Biophys. J.* **2003**, *84*, 3583–3593.
- (17) He, J.; Zhang, Z.; Shi, Y.; Liu, H. *J. Chem. Phys.* **2003**, *119*, 4005–4017.

- (18) Haliloglu, T.; Bahar, I.; Erman, B. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.
- (19) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505–515.
- (20) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2005**, *426*, 168–175.
- (21) Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110.
- (22) VandeVondele, J.; Röthlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 203–208.
- (23) Rosso, L.; Minary, P.; Zhu, Z. W.; Tuckerman, M. E. *J. Chem. Phys.* **2002**, *116*, 4389.
- (24) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (25) Bonomi, M.; Parrinello, M. *Phys. Rev. Lett.* **2010**, *104*, 190601.
- (26) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.
- (27) Redner, R.; Walker, H. *SIAM Rev.* **1984**, *26*, 195–239.
- (28) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584–590.
- (29) Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173–185.
- (30) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412–425.
- (31) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12347–12352.
- (32) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.
- (33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331–342.
- (34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (35) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (36) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- (37) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657–661.
- (38) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.