

Spatial Averaging: Sampling Enhancement for Exploring Configurational Space of Atomic Clusters and Biomolecules

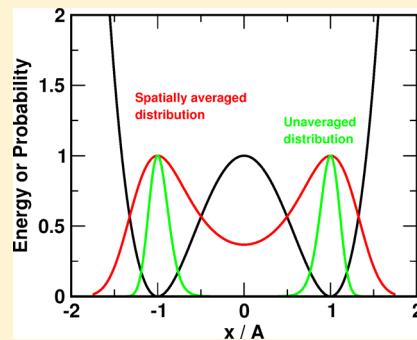
Florent Hédin,[†] Nuria Plattner,[‡] J. D. Doll,[§] and Markus Meuwly*,^{†,§}

[†]Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

[‡]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

[§]Department of Chemistry, Brown University, Providence, Rhode Island 02912, United States

ABSTRACT: Spatial averaging Monte Carlo (SA-MC) is an efficient algorithm dedicated to the study of rare-event problems. At the heart of this method is the realization that from the equilibrium density a related, modified probability density can be constructed through a suitable transformation. This new density is more highly connected than the original density, which increases the probability for transitions between neighboring states, which in turn speeds up the sampling. The first successful investigations included the diffusion of small molecules in condensed phase environments and characterization of the metastable states of the migration of the CO ligand in myoglobin. In the present work, a general and robust implementation including rotational and torsional moves in the CHARMM molecular modeling software is introduced. Also, a procedure to estimate unbiased properties is proposed in order to compute thermodynamic observables. These procedures are suitable to study a range of topical systems including Lennard-Jones clusters of different sizes and the blocked alanine dipeptide (Ala_2) in implicit and explicit solvent. In all cases, SA-MC is found to outperform standard Metropolis simulations in sampling configurational space at little extra computational expense. The results for (Ala_2) in explicit solvent are in good agreement with previous umbrella sampling simulations.



1. INTRODUCTION

Monte Carlo (MC) methods¹ are widely used in modern computer simulations to study high-dimensional, many-body systems.² One of their key features is their dimensional tolerance that makes it possible to study large systems with a significant number of degrees of freedom. Furthermore, when applied to atomic systems, and by choosing an appropriate statistical mechanical ensemble, MC simulations are useful in estimating the partition function, from which thermodynamic properties can be determined.

Despite their general utility, MC methods have practical limitations, one of which is related to rare-event sampling, which is a particular challenge.³ Conventional stochastic methods typically use random walk procedures for generating a statistical sampling of the desired equilibrium probability distribution, useful for obtaining numerical estimates. For systems in which configuration space is well connected, standard techniques such as the Metropolis–Hastings approach^{4,5} are efficient. However, often configuration space decomposes into poorly connected subregions, which makes realistic and exhaustive sampling problematic, and sampling needs to be enhanced. Several strategies have been developed in the past to address the rare event sampling problem, including parallel tempering (PT),⁶ umbrella sampling (US),⁷ metadynamics,⁸ or replica exchange (RE).⁹ These techniques either use a bias to drive the system from one region in configuration space to another, neighboring region (US, metadynamics), whereas PT and RE—which are related to

each other—expand thermodynamic state space. A broader overview of these techniques has been presented recently in the literature.^{2,3} Broadly speaking, the available techniques fall in one of the three following categories:

- (i) Trial move optimizations, as the displacement vector MC technique,¹⁰ or more recent studies specifically aiming at MC simulations of proteins,¹¹ but this kind of parameter tuning requires some *a priori* knowledge about the “shape” of the underlying potential energy surface.
- (ii) Parallel tempering,⁶ replica exchange,¹² and infinite swapping methods,^{13–17} which are based on repeated information exchange between copies of the simulation system, which are run at different values of an external control parameter (such as temperature). Lower-temperature replicas are enriched with knowledge coming from higher-temperature ones where high-energy barriers are more easily crossed. Several strategies for defining the tempering ensemble have been discussed.^{18–21}
- (iii) Through the addition of an external bias, such as a supplementary potential for filling basins of energy surface (metadynamics,⁸ flooding^{22,23}), auxiliary probability density (Tsallis weight sampling),^{24–26} energy smoothing methods,^{27–30} or constrained geometry (umbrella sampling).⁷

Received: June 18, 2014

Published: August 29, 2014

Spatial averaging MC (SA-MC) sampling belongs to this last category, where a new family of probability density functions are constructed.³¹ Until now, SA-MC has been applied to model systems and in special applications^{31,32} such as the diffusion of small molecules in condensed phase environments. The aim of the present work is (i) to introduce a general and robust implementation of SA-MC into CHARMM;³³ (ii) to generalize the available move set to include rotations and torsions; (iii) to investigate the possibility of determining unbiased thermodynamic properties from SA-MC in order to extract approximate thermodynamic information from the simulations; (iv) to apply SA-MC to the well-known problem of finding the optimal geometry of Lennard-Jones clusters (it is of particular interest to compare the efficiency in terms of the number of MC-steps compared to Metropolis sampling and the relative CPU requirements of the two approaches); and (v) to apply SA-MC to the conformational sampling of the blocked alanine dipeptide in implicit and explicit solvent.

2. COMPUTATIONAL METHODS

2.1. Spatial Averaging MC. In the canonical (NVT) ensemble, the probability $\rho(\mathbf{X})$ of observing a given system in state \mathbf{X} is related to its energy $V(\mathbf{X})$ through

$$\rho(\mathbf{X}) = \frac{1}{Z} e^{-\beta V(\mathbf{X})} \quad (1)$$

where $\mathbf{X} = X_1, \dots, X_k$ is a k -dimensional vector of coordinates (where $k = 3$ for MC or $k = 6$ for MD), populating a subset D of the configuration space \mathbb{R}^{kN} , Z is the canonical partition function $Z = \int_{D \subset \mathbb{R}^{kN}} e^{-\beta V(\mathbf{X})} d\mathbf{X}$, and $\beta = 1/k_B T$ is the inverse temperature and k_B the Boltzmann constant.

Monte Carlo (MC) methods¹ are one powerful way for sampling the high dimensional integral Z which runs over $3N$ degrees of freedom for a general Euclidean 3-space and for an N -particle system. The Metropolis–Hastings approach was specifically designed for addressing this problem when considering the canonical ensemble. Initially proposed for sampling the Boltzmann distribution,⁴ it was later extended to nearly all sampling problems.⁵ In practice, a system \mathbf{X} is stochastically modified leading to a new configuration \mathbf{Y} . Based on the energy difference $\Delta E = V(\mathbf{Y}) - V(\mathbf{X})$ the probability of accepting the new configuration is then

$$P_{\text{acc}} = \min\{1, e^{-\beta \Delta E}\} \quad (2)$$

For high energy barriers, the term $e^{-\beta \Delta E}$ is close to zero, and the probability of accepting such a move is extremely low. Previously introduced methods (PT/RE, US, metadynamics) addressed this problem by proposing a physical modification of the system (e.g., a set of temperatures for PT/RE). With SA-MC, increased sampling is achieved by directly modifying the underlying probability density function.^{31,32} In a one-dimensional notation, if the density to be sampled is $\rho(x)$, a new set of modified densities is obtained by writing

$$\rho(x, \epsilon) = \int_D P_\epsilon(y) \exp(-\beta V(x + y)) dy \quad (3)$$

where $P_\epsilon(y)$ is a normalized probability distribution with characteristic length scale ϵ . The parametrization of $P_\epsilon(y)$ is that of a Gaussian distribution with standard deviation ϵ . Adjusting this parameter allows to adapt the biasing distribution to the particular problem of interest. In practice, the convolution of the true distribution with $P_\epsilon(y)$ will decrease

the barriers of $V(x)$ and hence accelerate sampling of neighboring minima if ϵ is appropriately chosen. Furthermore, the Gaussian transform of the potential is centered around $\rho(x)$ so the integrals of the original and the transformed density are equal

$$\int_D \rho(x) dx = \int_D \rho(x, \epsilon) dx \quad (4)$$

Equation 4 is key to SA-MC, as it implies that thermodynamic properties derived from the modified density are related to those corresponding to the original density $\rho(x)$ for a given temperature. Let $\langle f(x) \rangle_0$ be a thermodynamic property (where the subscript 0 denotes an unbiased value) estimated through an average of the form

$$\langle f(x) \rangle_0 = \frac{\int_D \rho(x) f(x) dx}{\int_D \rho(x) dx} \quad (5)$$

By combining eqs 4 and 5, this average can be expressed by using the modified densities:

$$\langle f(x) \rangle_0 = \frac{\int_D \rho(x, \epsilon) \left(\frac{\rho(x)}{\rho(x, \epsilon)} f(x) \right) dx}{\int_D \rho(x, \epsilon) dx}$$

which can be simplified to

$$\langle f(x) \rangle_0 = \left\langle \left(\frac{\rho(x)}{\rho(x, \epsilon)} f(x) \right) \right\rangle_\epsilon \quad (6)$$

Hence, $\langle f(x) \rangle_0$ is expressed as an accumulated average of the instantaneous value $f(x)$ weighted by the ratio between the original and spatially averaged densities. Hence, the unbiased thermodynamic property of interest can be estimated.

As an example the Helmholtz Free Energy F as a thermodynamic function of state (ensemble NVT) is considered. The unbiased value F_0 estimated from a SA-MC simulation is

$$F_0 = F_\epsilon \frac{\rho(x)}{\rho(x, \epsilon)} \quad (7)$$

where F_ϵ is a biased estimate of F . In practice, the value of F for a given configuration x is estimated by counting the number of occurrences n of the configuration over all the sampled configurations N , which yields $\beta F = -\ln(n/N)$. By introducing a reference value F^0 for the free energy (for example the most stable configuration sampled), and by choosing a correct metric or reduced coordinates, it is possible to generate a surface of $\Delta F = F - F^0$: such energy landscapes are a powerful way of visualizing the configuration space and particularly useful for localizing minima regions, barriers, and saddle points.³⁴

As usual, when providing an estimate of any property it is important to also quantify the underlying statistical error. According to eq 7 errors in the estimates originate from (i) the error on F when counting the configurations and denoted as $\sigma(0) = -k_B T (\sigma(\rho(x))/\rho(x))$, and (ii) the error in the unbiasing ratio, directly related to the statistical variance on the spatially averaged densities. This variance can be estimated according to³¹

$$\sigma^2(\epsilon) = \frac{\rho(x)}{\rho(x, \epsilon)} \left(\frac{\rho(x)}{\rho(x, \epsilon)} - 1 \right) F_\epsilon \quad (8)$$

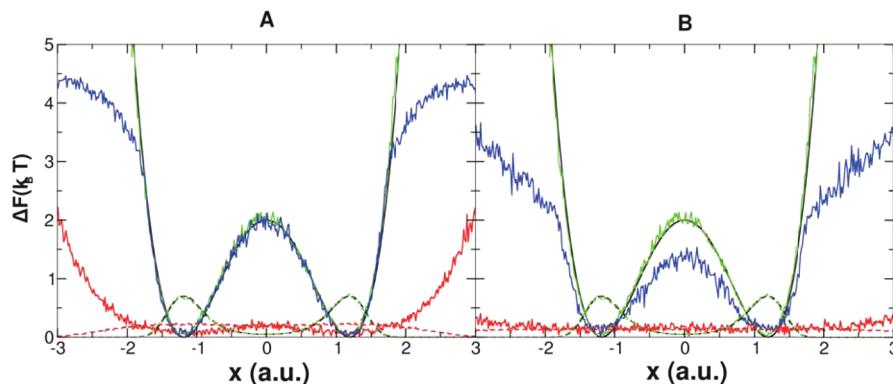


Figure 1. Reconstructed energy surface for the double well potential. Solid lines are for the surfaces ΔF and dashed lines are the corresponding densities $\rho(x)$. Left panel for ($W_e = 0.4$ and $N_e = 10$), right panel for ($W_e = 0.8$ and $N_e = 25$). Color code: analytical results (black), Metropolis MC (green), biased SA-MC (red), unbiased SA-MC (blue). a.u. = arbitrary units of distance.

The total error on the estimate of F_0 is

$$\sigma_{F_0} = \frac{1}{\sqrt{N}}(\sigma(0) + \sqrt{\sigma^2(\epsilon)}) \quad (9)$$

The $N^{-0.5}$ dependency in eq 9 is inherent to stochastic sampling methods.³⁵ However, by dividing the data in k data sets of a given size M (with $N = kM$) and by averaging over such blocks, the error can be reduced. More precisely, bootstrapping^{36–38} whereby only part of the data—randomly chosen from the overall distribution sampled by these four simulations—will be employed to estimate the error in the free energy profiles.

2.2. Algorithm and Implementation into CHARMM. The extension of SA-MC to multidimensional molecular systems has been successfully applied to the diffusion of small molecules (H_2 and CO) in condensed phase environments.³² This first algorithmic implementation was limited to translational and rotational moves, which makes possible to study diffusion processes. This will be generalized in the present work to also allow treatment of the configurational space of systems such as peptides and proteins.

The MC module³⁹ in CHARMM³³ is suitable for such an implementation as it allows the user to define an arbitrary set of moves for optimizing the sampling of a given molecular system. The main types of moves are (i) rigid translations of one or more atoms (RTRN), (ii) rigid rotations of one or more atoms around a center of rotation: this center may be another set consisting of one or more atoms, or the center of mass of the rotating atoms (RROT), (iii) dihedral angles torsions (TORS), and (iv) concerted rotations of dihedral angles (CROT). The current implementation handles (i–iii) in the NVT ensemble in explicit or implicit solvent. The present simulations were carried out with both the Analytical Continuum Electrostatics (ACE)^{40,41} implicit water model and the TIP3P⁴² explicit water model.

Starting from a trial configuration \vec{x}_0 of the system, a Gaussian distribution for M_e sets of N_e configurations with standard deviation W_e , centered around \vec{x}_0 is generated in SA-MC.^{32,43} The chosen MC move—such as translation or rotation—is then applied to all $M_e * N_e$ configurations and the corresponding energies $E_{\text{new}}^{(m,n)}$ are determined. Two sets of Boltzmann weights are then computed, one for the old and one for the new configurations: $E_{\text{old,Boltz}}^{(m,n)} = e^{-\beta E_{\text{old}}^{(m,n)}}$ and $E_{\text{new,Boltz}}^{(m,n)} = e^{-\beta E_{\text{new}}^{(m,n)}}$. For each set M_e , the difference between the aggregated

old and new weights is determined: $\delta_m = \ln(S_{\text{new}}^m / S_{\text{old}}^m)$ where $S^m = \sum^{N_e} E_{\text{Boltz}}^{(m,n)}$. Adding up all the δ_m yields $\delta = (1/M_e) \sum^{M_e} \delta_m$ from which also a variance $\sigma^2 = (1/M_e(M_e - 1)) \sum^{M_e} (\delta_m - \delta)^2$ can be computed. These quantities are then used for a modified acceptance/rejection criterion $\xi < \exp(-\beta(\delta + (\sigma^2/2)))$ (see eq 2).

For each MC move type, the new configurations have to be generated in the corresponding configurational space, such as for angle moves in the angular space around the initial \vec{x}_0 . This is accomplished as follows. (i) For a rigid translation, the procedure consists of adding a Gaussian distributed random number with zero mean and standard deviation W_e to the coordinates of the atoms that were selected for a particular move. (ii) For a rotation of a group of several atoms with coordinates X , a random angle θ , normally distributed between $-\theta_{\max}$ and $+\theta_{\max}$ is generated and the corresponding new coordinates are $X' = RX$ where R is a rotation matrix. (iii) Dihedral angles, defined as the intersection of two planes formed by four atoms, are also altered by drawing from a normal distribution and again by finding the Euler rotation matrix for the set of all atoms which are involved in the dihedral angle.

The ratio $\rho(x)/\rho(x,\epsilon)$, as used in eqs 6 and 7, is required for determining unbiased thermodynamic properties and is optionally stored for each frame of the trajectory in a dedicated file. This data can then be used in postprocessing from which the unbiased free energy and other observables can be estimated.

3. APPLICATIONS

In the following sections, SA-MC is applied to a range of three typical rare-event sampling problems, and its efficiency is compared to reference simulations, including standard Metropolis sampling. First, the current implementation together with the unbiasing procedure is tested on the double well potential to obtain thermodynamic properties.³¹ In the second example, the minimum energy structures of Lennard-Jones clusters are considered with particular focus on how to rapidly find the lowest energy configuration of such systems. The third and final example is the study of the free energy landscape of the blocked alanine dipeptide, which highlights the efficiency of SA-MC. For the first two examples, the simulations are performed with a dedicated code specifically written for the application whereas the third system is studied with the generalized CHARMM implementation described above.

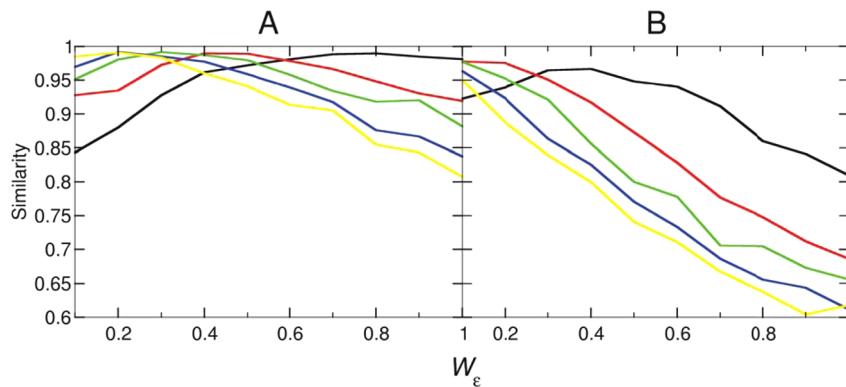


Figure 2. Similarity between reconstructed and theoretical surface for (a) barrier of $2k_B T$ or (b) barrier of $5k_B T$. Color code: $N_e = 5$ (black), $N_e = 10$ (red), $N_e = 15$ (green), $N_e = 20$ (blue), $N_e = 25$ (yellow).

3.1. Double Well Potential. To illustrate the efficiency of SA-MC but also the need for unbiasing when estimating thermodynamic properties, first a one-dimensional problem involving a simple double-well potential is studied. Explicitly, $V(x) = (x^2 - \sqrt{\lambda})^2$, where λ is the height of the barrier separating the two minima, which are located at $\pm(\lambda)^{1/4}$. Reduced units are used throughout which makes temperature dimensionless and energies are given in units of $k_B T$.

For a given temperature, the probability density of sampling x is $\rho(x) \propto \exp(-\beta V(x))$. Sampling $V(x)$ is sufficiently straightforward for low barriers that conventional MC yields the correct free energy profile. Therefore, the sensitivity of SA-MC to various choices of W_e and N_e ($M_e = 1$ in the present application) can be tested. For a reduced temperature of $\beta = 0.75$, 10^6 MC steps, and barrier heights between 1 and 10 simulations were carried out by using conventional MC and SA-MC. For the latter, $0.1 \leq W_e \leq 1.0$ in increments of 0.1 and $5 \leq N_e \leq 25$ in increments of 5.

The free energy curves are reconstructed and unbiased as explained in the computational methods part. For quantifying the similarity between the sampled density $\rho^\alpha(x)$ and the true normalized density $\rho(x)$ a score S^α is introduced:

$$S^\alpha = \frac{\int_{-\infty}^{+\infty} \rho(x) \rho^\alpha(x) dx}{\int_{-\infty}^{+\infty} \rho(x) \rho(x) dx} \quad (10)$$

where $\alpha = \text{MC}$ or SA-MC , respectively, and $\rho(x)$ is the true, normalized density. Hence, S^α measures the overlap between the sampled densities and the true Boltzmann density. For perfect sampling one should find $S = 1$.

Figure 1a is an example of reconstructing the FES for a barrier height of $\Delta F = 2k_B T$, where the theoretical surface $V(x)$ and the results of the Metropolis sampling (which overlaps ideally) are presented both with results from SA-MC with $W_e = 0.4$ and $N_e = 10$. Although SA-MC itself only poorly samples the reference FES, unbiasing as discussed in the Methods yields a very realistic FES (compare black and blue traces). Changing the parameters to $W_e = 0.8$ and $N_e = 25$ (Figure 1b) leads to almost uniform sampling with SA-MC (red trace). Despite this, the reconstructed, unbiased FES can capture the shape of the true FES although the free energy barrier is underestimated. This already highlights that SA-MC can be effectively used—even with unoptimized parameters W_e and N_e —to characterize the true *shape* of the free energy surface although barrier heights may only be qualitatively correct.

In a next step, the reconstructed (unbiased) FESs from SA-MC are further characterized, in particular with regards to the parameters W_e and N_e . For example, if the width W_e is too large, all information about the existence of local minima is washed out. Such considerations are of particular importance when applying SA-MC to a problem for which the underlying FES is incompletely or poorly characterized, that is, in cases where the positions and relative stabilizations of the minima are largely unknown. Figure 2a reports the similarity (estimated by using eq 10) between the reference and the unbiased SA-MC FES for barrier height $\lambda = 2$, $0.2 \leq W_e \leq 1.0$ and $5 \leq N_e \leq 25$. For the present case, increasing W_e improves the results initially for most N_e . However, beyond $W_e = 0.4$, the overlap between the reference and the SA-MC FES deteriorates. Hence, the sampling becomes less reliable. This is even more so for a larger barrier ($\lambda = 5$, panel b) for which small values of W_e give the best results.

This finding can be interpreted as follows. W_e is the width of the Gaussian distribution, that is, how far from the original configuration a new one will be generated, whereas N_e is the number of those additional configurations. Increasing both parameters increases the number of configurations generated, which are more and more distant from the original one, resulting in a large variance which causes an inaccurate estimate of the free energy. This is illustrated in Figure 3, where the free

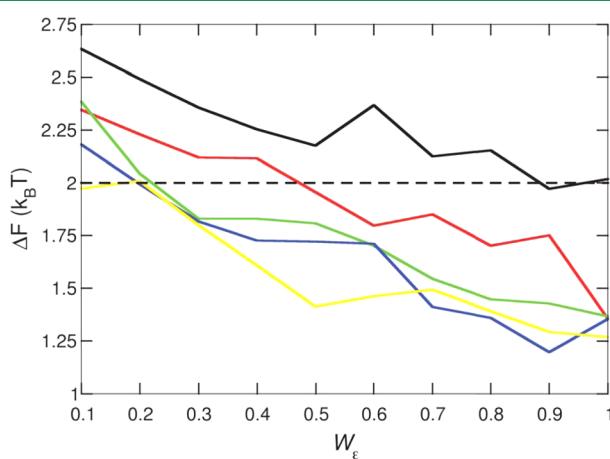


Figure 3. Unbiased barrier energy ΔF (reference value is $\Delta F = 2k_B T$, dashed black line) as a function of W_e . Systematic errors are of the order of $k_B T/100$. Color code (plain lines): $N_e = 5$ (black), $N_e = 10$ (red), $N_e = 15$ (green), $N_e = 20$ (blue), $N_e = 25$ (yellow).

energy at the top of the barrier ($\Delta F(x=0) = 2k_B T$) is reported for different sets of parameters N_e and W_e . For small values of N_e it is necessary to increase W_e for obtaining the correct value of $\Delta F = 2k_B T$. With larger N_e , a value of $W_e = 0.2$ is sufficient, and further increasing the Gaussian width will result in a less accurate value for $\Delta F(x=0)$.

In this first application, it is found that the bias introduced by SA-MC is a powerful feature that can more readily connect densities in local minima, separated by a barrier which is difficult to overcome with standard MC sampling. Furthermore, it is shown that the bias can be accounted for over a certain system parameter space (W_e and N_e) to faithfully reconstruct the true, underlying free energy profile. The degree to which this is possible depends on the system and the parameters chosen.

3.2. Global Minima of Lennard-Jones Clusters.

Lennard-Jones (LJ) clusters are an ideal class of systems to which MC-based sampling approaches can be applied. Some of the problems can be exhaustively sampled whereas others are computationally too demanding for this. Here, SA-MC is applied to determine low-energy configurations of LJ clusters of different sizes. The particular focus for this example is (i) whether or not the global minimum as known from the literature is found at all and (ii) the speed with which the global minimum is found. This also motivates the comparison of conventional MC and SA-MC in the present context. However, it should be mentioned that more established algorithms exist for global optimization.⁴⁴ LJ clusters are an ensemble of nonreactive atoms in vacuum (for example noble gases), interacting only through Lennard-Jones⁴⁵ potentials

$$V^{\text{LJ}} = 4\epsilon \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (11)$$

Here, r_{ij} is the distance between atoms i and j , ϵ is the depth of the potential well, and σ the distance at which $V^{\text{LJ}} = 0$. Again, reduced units are employed, that is, $\epsilon = \sigma = 1$, and the energy will be reported in units of ϵ . Extensive previous literature on these systems is available, and a Web site⁴⁶ provides a collection of known structures, lowest minima, and symmetry groups for clusters ranging from 2 to 1610 atoms: several MC methods,^{47–51} quantum calculations,⁵² MD simulations,^{53–55} parallel tempering,^{56,57} or others such as discrete path sampling^{58,59} were used for characterizing the systems, and LJ_N (with N the number of atoms) clusters became reference systems for methods dedicated to finding global minima. The number of local minima grows exponentially as a function of N , and hence, determining the global minimum of such clusters is a computationally challenging problem. As an example, between $N = 2$ and $N = 33$ the number of known minima increases from 1 to $\approx 4 \times 10^{14}$. Nevertheless, some recent studies were able to treat the broken ergodicity and then provide the correct number of minima for the LJ_{31} and LJ_{75} clusters.⁶⁰

The low energy minima for various LJ_N clusters were investigated by both, conventional MC and SA-MC. Specifically, the systems included $N = 13, 19, 31, 38, 55$, and 75. Some of the systems are relatively “easy” while others—such as LJ_{38} , see Figure 4—are known to be very challenging (see below). The methodology applied for all coming examples is as follows: (i) 10^4 independent runs are started from the same initial (random) configuration. (ii) At the end of each step, if the

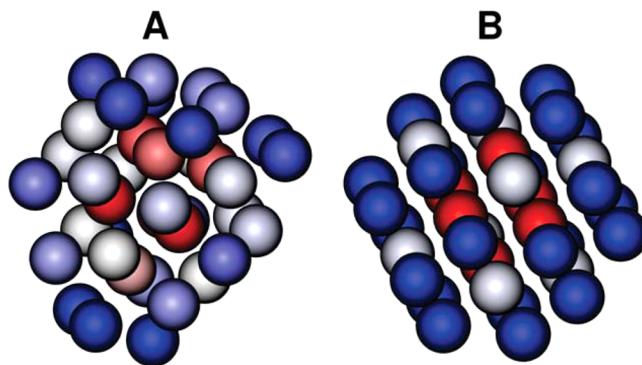


Figure 4. Lowest energy configurations found for the LJ_{38} atoms, obtained with SA-MC. The structure in panel a has an energy of $E = -171.357\epsilon$ (see Table 1) whereas the structure in panel b is that of the absolute minimum ($E = -173.928\epsilon$) found when starting from the cluster LJ_{37} and randomly adding an atom. Red atoms are closer to the center of mass of the cluster than gray and blue ones.

energy difference relative to the reference configuration is less than 5ϵ the system is minimized, and if the known lowest energy minimum structure is obtained (tolerance of $10^{-4} \times \epsilon$) the calculation is stopped and considered as converged; otherwise, the simulation continues. (iii) If the global minimum is not reached after a given number of steps (depending of cluster size) the simulation is considered to be not converged.

For LJ_{13} , the global minimum has an energy of $E = -44.327\epsilon$ (see Table 1). Figure 5 shows a cumulative distribution of the required number of steps before reaching the global minimum for conventional MC and SA-MC with different parameters [$W_e; M_e; N_e$]. After the 10^6 MC steps considered here, only 24% of the MC simulations are able to locate the global minimum energy structure. This compares with between 50% and 98% for SA-MC, depending on the choice of M_e and N_e . In general, SA-MC outperforms conventional MC considerably. For $W_e = 0.25$ and 0.5 a clear improvement is observed, as almost all simulations converge before 2.5×10^5 steps. Results are particularly noteworthy with $W_e = 0.5$ for which 98% of the simulations converged during the first 10^5 steps. For $W_e = 1.0$ the convergence speed slows down. One explanation is that depending on the value for W_e —typically the larger W_e the flatter the FES—the SA-MC-modified densities become too connected which changes the topology of the FES such as to slow down convergence. Nevertheless, this may be corrected by using increased values of M_e and N_e , which leads to variance reduction. However, the computational time would also increase.

The previous conclusions are supported by an analysis of the median of the number of steps for converged simulations, that is, the value for which 50% of the calculations reach the minimum energy structure. For conventional MC, this value is 3.3×10^5 compared to 6×10^4 , 2×10^4 , and 3×10^5 for SA-MC with [0.25;5;5], [0.5;5;5], and [1.0;5;5], respectively. Hence, for the best performing SA-MC simulation, the average number of steps required to reach the global minimum is smaller by a factor of 30 compared to conventional MC. It is also possible to determine the rate at which the various simulations converge by fitting the cumulative successful runs to an empirical relationship $y = d \tanh((ax + b)/d)$ where d describes the asymptotic convergence (plateau of the number of converged simulations, ideally 10000) and a describes the growth of the first part of the curve (i.e., how rapidly the

Table 1. Minimum Energy Configurations (in Units of ϵ) for All LJ Clusters Studied, and Best Convergence Rates Observed, for MC and SA-MC^a

LJ _N	E_{ref}^{47}	E_{MC}	$E_{\text{SA-MC}}$	steps	conv. MC (%)	conv. SA-MC (%)
13	-44.326	-44.326	-44.326	10^6	24	98
19	-75.659	-75.659	-75.659	10^6	22	97
31	-133.586	-126.081	-133.586	10^8		26
38(a)	-173.928	-160.556	-171.357	10^8		
38(b)	-173.928	-173.928	-173.928	10^9	3	35
37 + 1	-173.928	-170.807	-173.928	5×10^7		6
55	-279.248	-279.248	-279.248	10^8	16	65
75	-397.492	-381.173	-397.492	10^8		28

^aReference values are from the literature.⁴⁷ Numbers in bold face are unconverged values. The “steps” column indicates how long were both MC and SA-MC simulations. What differs between 38(a) and 38(b) is the number of steps. 37 + 1 means that the starting point is the lowest energy geometry of LJ₃₇ to which a 38th atom is included.

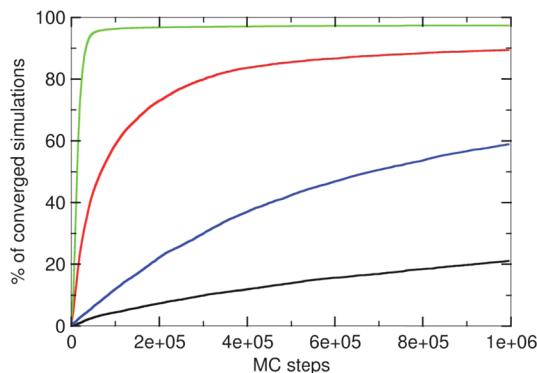


Figure 5. Convergence analysis for MC (black), and SA-MC simulations with different sets of parameters [$W_e; M_e; N_e$]: [0.25;5;5] (red), [0.5;5;5] (green), and [1.0;5;5] (blue).

plateau is reached). The parameter b ensures that the fit passes through the origin. For conventional MC, $a = 1.2$ compared with 5.6, 48.2, and 1.6 from SA-MC, which quantifies the above observations about the median. For parameter d , the fit yields 2400, 8700, 9700, and 6200 for the four simulations. This, together with the observations for parameter a suggests that the rate of successful runs for the worst SA-MC simulation is still better than that of conventional MC whereas the number of successful runs is larger by a factor of 3. On the other hand, the best performing SA-MC simulation is about 30 times as efficient while finding at the same time the global minimum in almost all simulations (98%).

The computational overhead in using SA-MC is in the $M_e \times N_e$ additional energy evaluations which, however, can be easily parallelized. In the present case, a factor of $5 \times 5 = 25$ is expected for a given number of MC steps (here 10^6). If all 10^4 simulations are run for 10^6 MC steps, SA-MC with [0.5;5;5] is 23-times slower than conventional MC. However, if simulations are terminated when the lowest minimum is found, this reduces to a factor of 1.2. Hence, in cases where suitable termination criteria can be found, the computational overhead of SA-MC is well below an order of magnitude compared to conventional MC with the added value of the much increased likelihood for locating the correct lowest energy configuration.

For the larger LJ clusters, Metropolis MC simulations have difficulties in successfully locating the known minima at all. In order to assess the performance of SA-MC for such cases, additional simulations were carried out for LJ₁₉, LJ₃₁, LJ₃₈, LJ₅₅, and LJ₇₅. The same procedure as before is used except for the total number of MC or SA-MC steps, which was increased to

10^9 for larger clusters. For LJ₁₉, the convergence speed analysis gives results similar to those presented in Figure 5; that is, SA-MC reaches the global minimum (energy -75.659ϵ , see Table 1) in much fewer steps than regular Metropolis, when using 10^6 steps. The LJ₅₅ and LJ₇₅ are much larger systems and Metropolis sampling is extremely slow to obtain converged results. Table 1 shows that for LJ₅₅ MC and SA-MC converge to the reference value from the literature, the convergence rate for MC is 15.8% with a median number step required of 7.6×10^7 (the maximal number of steps being 10^8), and for SA-MC the numbers are 65% and 5.4×10^6 , respectively, that is, one order of magnitude faster when just considering the number of steps. For LJ₇₅, conventional MC sampling is unable to locate the global minimum within 10^8 steps. On the contrary, SA-MC does find this minimum for 28% of the simulations within a median number of steps of 5.1×10^7 .

The LJ₃₁ and LJ₃₈ clusters are known for their funneled energy landscape.^{47,49,58,61,62} LJ₃₈ is a particularly interesting system as it has a double-funnel landscape, one ending in the global minimum, the other in the second minimum. Doye et al. showed with disconnectivity graphs^{49,50} that 446 minima are related to the second funnel but only 28 to the first one, making the transition from one funnel to the other extremely rare. With 10^8 MC steps, our implementation of the Metropolis algorithm was unable to converge to the lowest known minimum for both clusters, which are at -133.586ϵ and -173.928ϵ , respectively, see Table 1. The best configurations sampled in this set of simulations are still 6ϵ and 13ϵ higher in energy than the known minima. Contrary to that, SA-MC successfully converged for the LJ₃₁ cluster (see Table 1) with similar sets of parameters as for LJ₁₉, but for LJ₃₈ (Figure 4a) the best energy obtained is still 2.5ϵ too high (-171.357ϵ compared to -173.928ϵ , cf. Table 1). A second set of 10 000 simulations for the LJ₃₈ cluster was carried out with 10 times more Monte Carlo steps (10^9 instead of 10^8 , see Table 1 line 38(b)). In this case, both MC and SA-MC find the known minimum energy structure^{47,49,61} for 3 and 35% of the simulations, respectively.

A final set of 10 000 simulations for the LJ₃₈ cluster was carried out using a slightly different approach: instead of starting from a fully random initial configuration, the lowest minimum of the LJ₃₇ cluster (which was successfully found by SA-MC) was employed and randomly a 38th atom was added to the system. Then, simulations were run for 5×10^7 steps. The lowest energy obtained from the Metropolis algorithm is then -170.807ϵ , which is considerably closer to the best minimum with fewer MC steps (see 38(a) versus 37 + 1 in Table 1), and -173.928ϵ for SA-MC, which is the correct

minimum energy structure (Figure 4b). Overall, it is found that SA-MC successfully converges to the global minimum for all the studied LJ_N clusters, including both funneled clusters (LJ_{38}) and larger clusters such as LJ_{75} .

3.3. Blocked Alanine Dipeptide in Implicit and Explicit Water.

The blocked alanine dipeptide (Ac-Ala-N-H-Me, Figure 6) has been used as a test system for computational

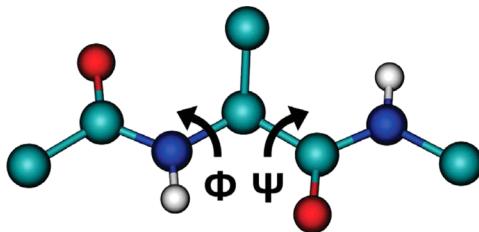


Figure 6. Blocked alanine dipeptide (Ac-Ala-N-H-Me), and the two dihedral angles of interest Φ ($C-N-C_\alpha-C$) and Ψ ($N-C_\alpha-C-N$).

studies^{63–79} of conformational equilibria and free energy landscape reconstruction and analysis. This dipeptide contains many of the structural features of proteins, including the two (ϕ, ψ) dihedrals angles, NH and CO groups capable of H-bond

formation, and a methyl group attached to the C_α atom. Successful studies used quantum chemistry, MD and MC simulations, and several conformations were identified:^{63,64,67,68,70,79} (i) β , also called C_5 , for $(\phi, \psi) \sim (-140^\circ, 150^\circ)$, (ii) $C_{7\text{eq}}$ for $(\phi, \psi \sim -90^\circ, 80^\circ)$, (iii) α_R (right-handed α helix) for $(\phi, \psi) \sim (-80^\circ, -60^\circ)$, (iv) α_L (left-handed α helix) for $(\phi, \psi) \sim (60^\circ, 60^\circ)$ and (v) $C_{7\text{ax}}$ for $(\phi, \psi) \sim (60^\circ, -60^\circ)$. One suitable way to visualize the free energy landscape for the conformations and the transitions between them is to report an energy surface as a Ramachandran plot.⁸⁰ Simulations for the blocked alanine dipeptide were carried out both in implicit (Analytical Continuum Electrostatics (ACE)^{40,41}) and explicit solvent (TIP3P⁴² water). ACE is known for providing a meaningful description of solvation effects for peptides.^{17,67,81,82} In the following, results from simulations with ACE are first described. Next, the simulations in explicit water are summarized.

Initially, two reference simulations were carried out. They included an MD and a Metropolis MC simulation and served as benchmarks with which to compare the SA-MC simulations. For the latter simulations with a range of parameters [$W_e; M_e; N_e$] were carried out. In all cases, the blocked alanine dipeptide is treated in a united atom representation (12 atoms, see Figure 6), nonbonded interactions are fully calculated, and

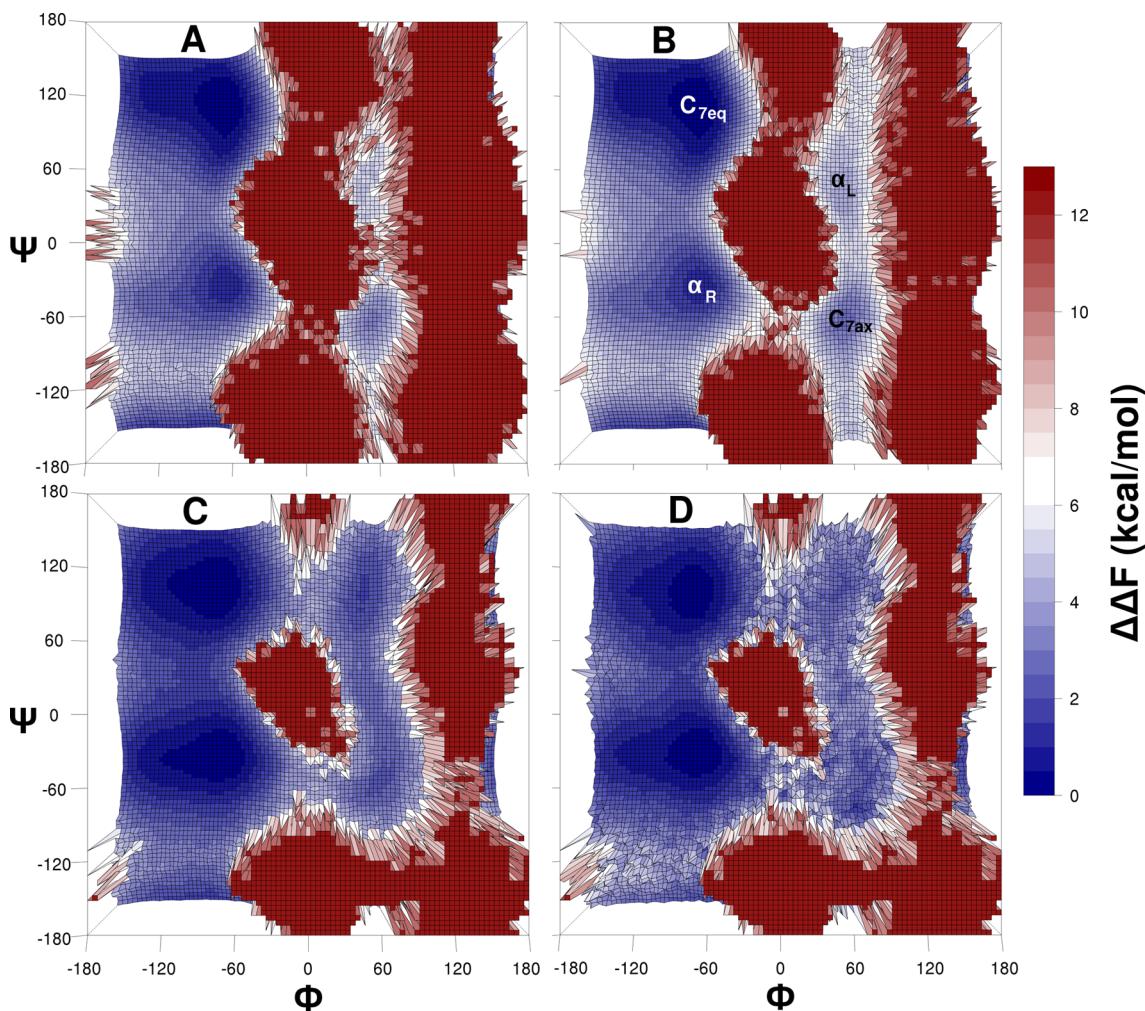


Figure 7. FES of alanine dipeptide: (A) MD, $1.5 \mu s$, 300 K; (B) Metropolis MC, 100×10^6 steps, 300 K; (C) biased SA-MC with parameters [0.5; 5; 5]; (D) unbiased SA-MC (same parameters), 5×10^6 steps, 300 K. All free energies are reported relative to the $C_{7\text{eq}}$ minimum.

the temperature is 300 K in the NVT ensemble. For simulations with the ACE implicit solvent, default parameters, such as Born solvation radii, dielectric constants, and atomic volumes, are taken from the literature.^{40,41} The MD simulations use the velocity Verlet integrator with the Nosé–Hoover thermostat for a simulation time of 1.5 μ s, a cutoff of 12 Å and a time step of $\Delta t = 0.5$ fs. The MC simulation was run for 10^8 steps. For SA-MC, simulations with several parameter sets were carried out: (i) $W_e \in \{0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0\}$, (ii) $M_e \in \{5, 10, 15, 20\}$, and (iii) $N_e \in \{5, 10, 15, 20\}$. M_e and $N_e = 5$ or 10 proved to be sufficient for the present purpose (there is no gain with larger parameters justifying the overhead). Here, results for $N_e = 5$ are presented.

Simulations in Implicit Solvent. Figure 7 shows the Helmholtz Free Energy Surfaces (FES) for both MD (A) and the Metropolis (B) simulations. It is first observed that both surfaces are quite similar to each other and closely resemble those obtained previously in the literature using the same computational setup.^{67,79} The data reported in Figures 7A and B already indicate that the barrier regions between the basins are not well sampled. This is true in particular for the MD simulations. The four following regions (see labels in Figure 7) are sampled sufficiently for providing an estimate of the associated free energy differences: (i) $C_{7\text{eq}}$ (top left basin of lowest energy), (ii) α_R (bottom left), (iii) $C_{7\text{ax}}$ (bottom right), and (iv) α_L (top right). Positions and estimates for the free energy for those four minima are summarized in Table 2. A

Table 2. Relative Free Energies (kcal/mol) and Minima Locations for the Blocked Alanine Dipeptide at 300 K, for MD, MC, SA-MC Simulations, and Three External References,^{64,67,79} All Using the ACE Implicit Solvent Model^a

basin	methods			position (Φ, Ψ)
	ΔF MD	ΔF MC	ΔF SA-MC	
$C_{7\text{eq}}$	0.00 ± 0.01	0.00 ± 0.02	0.00 ± 0.07	($-83^\circ, 136^\circ$)
α_R	1.10 ± 0.03	1.16 ± 0.04	0.21 ± 0.06	($-79^\circ, -42^\circ$)
$C_{7\text{ax}}$	3.26 ± 0.16	2.91 ± 0.12	3.11 ± 0.08	($67^\circ, -75^\circ$)
α_L	4.62 ± 0.32	4.86 ± 0.36	4.12 ± 0.09	($47^\circ, 55^\circ$)
references				
basin	ΔF ref 64	ΔF ref 67	ΔF ref 79	
$C_{7\text{eq}}$	0.00	0.00	0.00	
α_R	0.71	1.5	0.93	
$C_{7\text{ax}}$	4.34	4.1	2.94	
α_L	4.35	5.0	4.27	

^aThe statistical error was estimated using bootstrapping described previously, and \pm values represent a 95% confidence interval. All free energies are shifted relative to the $C_{7\text{eq}}$ structure which is the reference energy.

95% statistical confidence interval is provided (see previous description of the bootstrapping procedure) for MC, MD, and SA-MC simulation. The fact that this error is somewhat larger for SA-MC than for MC is caused by the additional error introduced by the unbiasing step of SA-MC (eq 8). Nevertheless, when considering higher energy minima as $C_{7\text{ax}}$ and α_L , this value is several times lower than the error estimated for the MD case, where the poor sampling causes an error of 0.32 kcal/mol. Furthermore, the highest error estimated for SA-MC is only 0.09 kcal/mol. It is also of interest to briefly comment on the effect of using bootstrapping for error

estimation. For example, directly using eq 9 without bootstrapping leads to an error of 0.18 kcal/mol for the α_L structure with SA-MC, which is reduced to 0.09 kcal/mol when using bootstrapping.

Figure 7C shows the FES from simulations with the SA-MC algorithm, with parameters $W_e = 0.5$ and $M_e = N_e = 5$ whereas panel D reports the unbiased FES from the same data. Compared to the MD and conventional MC simulations, SA-MC leads to a much improved sampling of the valley around $\Phi = 75^\circ$, and more specifically the two saddle points connecting the left and right parts of the FES. Such transitions are typically rare events in Metropolis MC but rather well sampled within SA-MC. Differences between the biased and unbiased SA-MC FESs are minor. On the biased FES (Figure 7 C), SA-MC lowers barriers by ≈ 1.2 kcal/mol, which means that the corresponding states are better sampled.

Numerical values for the relative free energy values of the minima and at the top of the barriers are summarized in Table 3. First, it is observed that the current MD and MC simulations

Table 3. Comparison of ΔF from (a) MD, (b) Unbiased Targeted MD Simulations,⁶⁴ (c) MC, and (d) SA-MC^a

	(a) MD				(b) ref 64			
	$C_{7\text{eq}}$	α_R	$C_{7\text{ax}}$	α_L	$C_{7\text{eq}}$	α_R	$C_{7\text{ax}}$	α_L
$C_{7\text{eq}}$	0.0	3.25			0.0	2.61		6.47
α_R	3.25	1.10			2.61	0.71	6.88	
$C_{7\text{ax}}$			3.26				6.88	4.34
α_L				4.62	6.47		5.98	4.35
(c) MC					(d) SA-MC			
	$C_{7\text{eq}}$	α_R	$C_{7\text{ax}}$	α_L	$C_{7\text{eq}}$	α_R	$C_{7\text{ax}}$	α_L
$C_{7\text{eq}}$	0.0	3.37			0.0	1.96		5.08
α_R	3.37	1.16			1.96	0.21	4.91	
$C_{7\text{ax}}$			2.91	5.66		4.91	3.11	4.20
α_L			5.66	4.86	5.08		4.20	4.12

^aDiagonal entries from Table 2 are stabilization energies relative to the global minimum $C_{7\text{eq}}$, whereas off-diagonal entries refer to the barriers between the minima. Empty cells indicate that the direct transition was not observed or is not possible. All free energies are reported relative to the $C_{7\text{eq}}$ structure, which is the reference energy.

only partially sample the FES compared to previous targeted MD simulations⁶⁴ and the SA-MC simulations. On the other hand, SA-MC and the reference simulations⁶⁴ sample similar amounts of the available configuration space. In general, the location of the minima and their energy is similar to that found from previous work.^{63,65,67,68} The $C_{7\text{eq}}$ minimum is the most stable state on the FES for all types of simulations, followed by α_R . Its relative stabilization energy compared to the global minimum is ≈ 0.2 kcal/mol from unbiased SA-MC, which compares with 0.7 kcal/mol⁶⁴ and above 1 kcal/mol from MD and MC simulations.

The relative stability of the $C_{7\text{ax}}$ and α_L structures from unbiased SA-MC are close to the MD simulations and differ by about 1 kcal/mol from reference simulations in the literature.^{64,67} This suggests, that SA-MC in the present case is a suitable method to locate stable and metastable states on the FES with high confidence but that the quality of the unbiasing depends somewhat on the state considered.

It is also interesting to consider the energy at the top of the barriers separating two stable conformations. This information is summarized in Table 3. In general, the unbiased SA-MC data follow those from previous simulations.⁶⁴ Typically, the

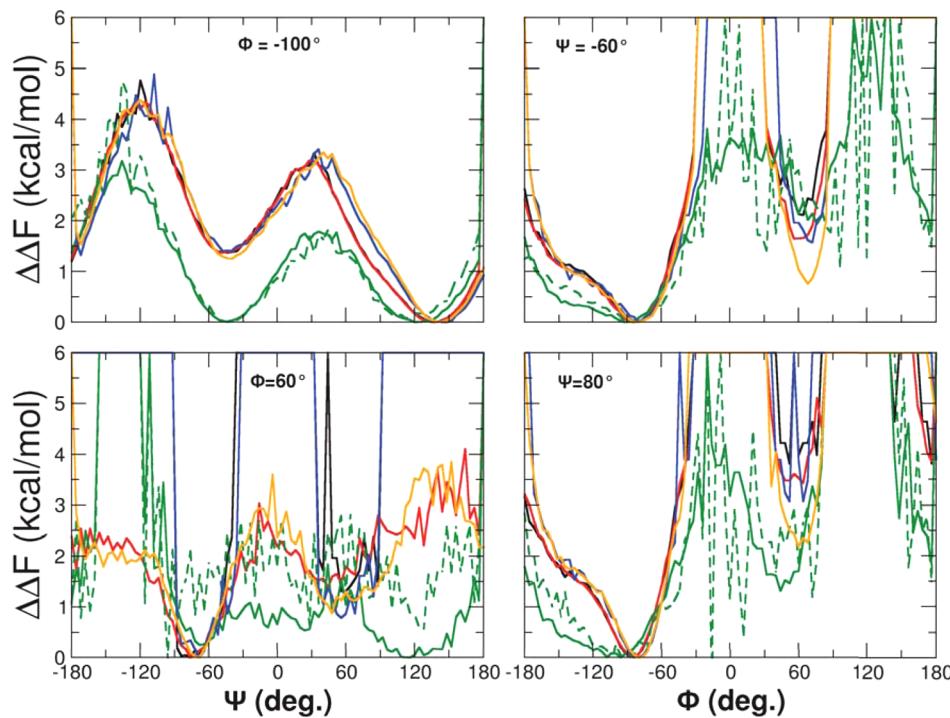


Figure 8. Slices through the FES from Figure 7 for MD (black), MC (red), SA-MC biased (dashed green), SA-MC unbiased (green), parallel tempering¹⁷ (blue), and infinite swapping¹⁷ (orange). SA-MC parameters are [0.5;5;5].

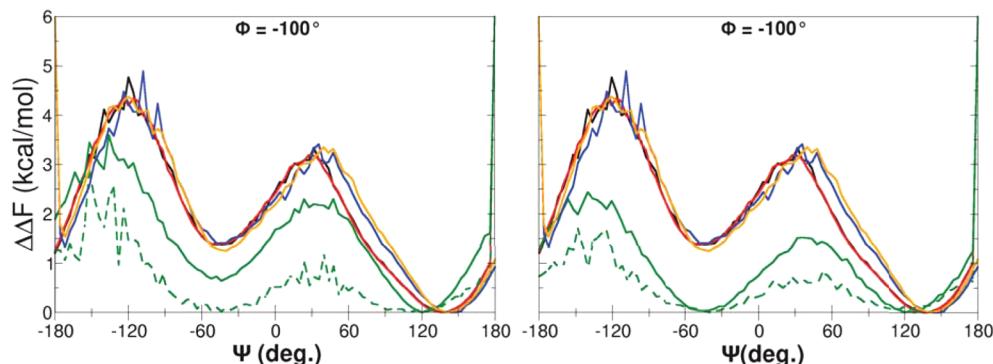


Figure 9. Slices through the FES, showing the influence of SA-MC parameters $[W_e;M_e;N_e]$ on the biased and unbiased SA-MC energy profiles. SA-MC parameters are [0.1;10;10] (left) and [0.1;10;15] (right). MD (black), MC (red), SA-MC biased (dashed green), SA-MC unbiased (green), parallel tempering¹⁷ (blue), and infinite swapping¹⁷ (orange) are shown as separate curves. Free energy is shifted in order to have a value of 0.0 for the C_{eq} minimum.

transition barriers are lower by about 1 kcal/mol but all orderings of the barriers agree with the data from the literature.

Figure 8 shows slices through the FESs from Figure 7 together with cuts from parallel tempering and infinite swapping simulations using the same setup of the systems,¹⁷ for $\Phi = -100^\circ$ and $\Psi = -60^\circ$, (left panels) and $\Psi = -60^\circ$ and $\Psi = 80^\circ$ (right panels). For the cut at $\phi = -100^\circ$, the topography of the FES from SA-MC is similar to all four other methods although quantitatively differences can be up to 1 kcal/mol for barriers and more for the secondary minimum. For the other three cuts, it is noted that there are much fewer unsampled regions (spikes) when using SA-MC than compared to any other method. Again, the unbiased SA-MC results underestimate the barriers and overstabilize the metastable states. However, from a sampling perspective SA-MC is clearly superior to MC: with 20 times fewer steps (5×10^6 for SA-MC against 100×10^6), for a similar CPU time usage, and transition

regions are considerably more sampled with SA-MC than for MD, MC, and PT simulations.

Figure 9 shows slices through the same FES as in Figure 8, that is, for $\phi = -100^\circ$, but reports results from simulations with different sets of SA-MC parameters: [0.1;10;10] (left) and [0.1;10;15] (right). It is apparent that the choice of SA-MC parameters influences the results. The data reported in Figure 9a better reproduces the reference simulations than the data in Figure 9b.

Simulations in Explicit Solvent. Sampling the free energy landscape of blocked alanine dipeptide in explicit water is computationally much more challenging.^{83–86} The present system consists of 462 water molecules to which SHAKE constraints⁸⁷ are applied and one blocked (Ala_2). The nonbonded cutoff parameter is 12 Å. Figure 10 reports the 2-dimensional FES obtained from 10^8 steps of SA-MC simulations with parameters [0.1;5;5] and Figure 11 reports

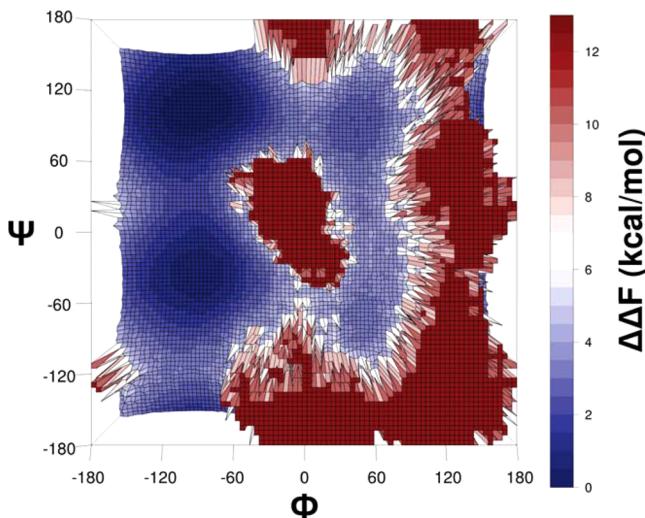


Figure 10. Unbiased FES for SA-MC from simulations of blocked alanine-dipeptide in explicit TIP3P water and with SA-MC parameters [0.1;5;5]. The number of steps is 100×10^6 steps. All free energies are reported relative to the $C_{7\text{eq}}$ minimum.

slices (for $\Phi = -100^\circ$ and $\Phi = -60^\circ$, (left panels) and $\Psi = -60^\circ$ and $\Psi = 80^\circ$ (right panels)) through the FES of Figure 10 for SA-MC. The overall topology of the FES in implicit and explicit solvent is similar. However, it is noted that the transition between $C_{7\text{eq}}$ and α_R is considerably wider in explicit water. Table 4 shows the energy estimated for the four known minima of the FES of Figure 10. Data from simulations with ACE are also included for comparison. The reference data is from refs 64 and 88, which was determined from MD simulations with both explicit solvent and a Generalized Born implicit solvent, and error bars represent a 95% confidence interval.

Table 4. Free Energies (in kcal/mol) Relative to the $C_{7\text{eq}}$ Minimum from SA-MC (ACE), SA-MC (TIP3P) Simulations Compared to Reference Data from the Literature, Which Employed Umbrella Sampling in Explicit Solvent^{64,88a}

basin	methods and references				
	SA-MC (ACE)	SA-MC (TIP3P)	ref 88	ref 64	position (Φ, Ψ)
$C_{7\text{eq}}$	0.00 ± 0.07	0.00 ± 0.08	0.00	0.0	$(-83^\circ, 136^\circ)$
α_R	0.21 ± 0.06	1.20 ± 0.07	1.30	1.41	$(-79^\circ, -42^\circ)$
$C_{7\text{ax}}$	3.11 ± 0.08	2.99 ± 0.08	NA	3.85	$(-67^\circ, -75^\circ)$
α_L	4.12 ± 0.09	3.60 ± 0.10	3.80	4.38	$(47^\circ, 55^\circ)$

^aFor $C_{7\text{ax}}$ no data are available from ref 88. The statistical error in the present work was estimated from bootstrapping and \pm values represent a 95% confidence interval.

It is found that the $C_{7\text{eq}}$ minimum is still the most stable state, followed by α_R , $C_{7\text{ax}}$, and α_L . However, the relative stabilizations are somewhat altered in that α_R is destabilized relative to $C_{7\text{eq}}$ whereas $C_{7\text{ax}}$ and α_L are somewhat stabilized. Comparison with literature data shows that for the simulations in explicit solvent the present results agree favorably for $C_{7\text{eq}}$, α_R , α_L , and for $C_{7\text{ax}}$ when available (see ref 64). It should be noted that no numerical values are provided in ref 88, and the numbers reported here have been inferred from the graphical illustrations (not possible for $C_{7\text{ax}}$). Comparison with ref 64, which also provided values for simulations with the ACE model, shows good agreement with results obtained with SA-MC; however, once again, it appears that the α_R minimum is somewhat overdestabilized when using SA-MC.

4. CONCLUSIONS AND OUTLOOK

In the present work, a practical and comprehensive implementation for spatial averaging MC (SA-MC) simulations into the CHARMM general purpose atomistic simulation

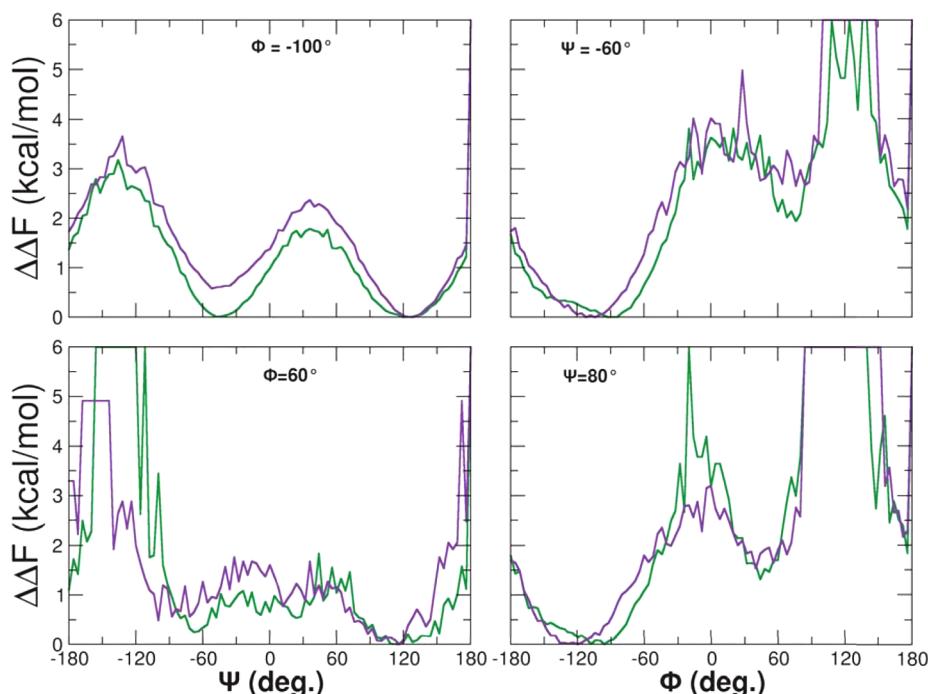


Figure 11. Slices through the FES from Figures 7 and 10: unbiased SA-MC with ACE (green), unbiased SA-MC with TIP3P (purple). All values are reported relative to the $C_{7\text{eq}}$ minimum.

program has been described. Also, an unbiasing procedure is discussed which allows to estimate thermodynamic observables. The implementation and unbiasing strategy are validated for model and topical systems including the double well potential, Lennard-Jones clusters and the blocked alanine dipeptide in implicit and explicit solvent. The considerably increased efficiency for exploring configuration space has been demonstrated for all three applications. However, the degree to which this is possible depends on the properties and connectivity of the systems' conformational space, which is usually a priori unknown. The central asset of SA-MC is that it generates a more highly connected ensemble, which makes exploration of the underlying free energy surface more readily possible.

It is expected that SA-MC can be beneficial for a range of future applications. As already indicated, SA-MC can be used to efficiently explore configurational space, based on which unbiased free energy surfaces can be obtained from the spatially averaged distribution. Furthermore, SA-MC is well suited to approximately locate transition states and to characterize the transition state ensemble.^{89,90} This is the starting point for enhanced exploration of barrier-crossing problems in more complex systems (such as small solvated peptides or proteins), which is typically difficult to achieve from standard MC or MD simulations. Given that SA-MC primarily connects neighboring metastable states, which are usually separated by barriers of a few $k_B T$, we expect SA-MC to perform well for such problems as was already demonstrated for the solvated dipeptide in the present work. Also, SA-MC can be employed to find approximate reaction coordinates, which is useful for subsequent umbrella sampling simulations.⁷ Finally, SA-MC could be employed together with Hamiltonian replica exchange molecular dynamics simulations (H-REMD).⁹¹ Hamiltonian replica exchange can be used for studying several types of problems, but in practice, its performance depends substantially on the details of the biased Hamiltonian. Similar to combining umbrella sampling simulations with H-REMD,⁹² employing SA-MC together with H-REMD could be potentially beneficial and provide a systematic way to generate biased Hamiltonians.

A common characteristic of all MC methods is that simulation parameters such as the move range, the acceptance ratio, or the swapping rate need to be optimized to some extent to obtain computational performance. This is also the case for SA-MC. One future improvement for the SA-MC algorithm is therefore to facilitate finding optimized sets of parameters [$W_e; M_e; N_e$] during the simulation. It is not necessary to use the same values for each of the MC steps because of the Markovianity of the procedure. The examples investigated here in more detail emphasize that larger values of the system parameters enhance the sampling of barriers and transition states at the cost of extra computational time. Hence, another possible improvement concerns the decrease of those parameters for regions well sampled by the MC algorithm for speeding up the sampling, and to increase them for poorly sampled regions, possibly during the simulation by using an "on-the-fly" optimization technique. This will be important for applying efficiently the SA-MC algorithm to larger systems.

AUTHOR INFORMATION

Corresponding Author

*Email: m.meuwly@unibas.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation through grants 200021-117810 and the NCCR MUST (to M.M.). J.D.D. acknowledges support from the National Science Foundation award DMS-1317199.

REFERENCES

- (1) Metropolis, N.; Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
- (2) Kroese, D.; Taimre, T.; Botev, Z. *Handbook of Monte Carlo Methods*; John Wiley & Sons, Inc.: Hoboken, NJ, 2011.
- (3) Rubino, G.; Tuffin, B. *Rare Event Simulation Using Monte Carlo Methods*; John Wiley & Sons, Inc.: Hoboken, NJ, 2009.
- (4) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087.
- (5) Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109.
- (6) Earl, D. J.; Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910.
- (7) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Chem. Phys.* **1977**, *23*, 187–199.
- (8) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (9) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and application to spin glass simulations. *J. Phys. Soc. Japan* **1996**, *65*, 1604–1608.
- (10) Voter, A. F. A Monte Carlo method for determining free-energy differences and transition state theory rate constants. *J. Chem. Phys.* **1985**, *82*, 1890.
- (11) Betancourt, M. R. Optimization of Monte Carlo trial moves for protein simulations. *J. Chem. Phys.* **2011**, *134*, 14104.
- (12) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607.
- (13) Plattner, N.; Doll, J. D.; Dupuis, P.; Wang, H.; Liu, Y.; Gubernatis, J. E. An infinite swapping approach to the rare-event sampling problem. *J. Chem. Phys.* **2011**, *135*, 134111.
- (14) Dupuis, P.; Liu, Y.; Plattner, N.; Doll, J. D. On the infinite swapping limit for parallel tempering. *Multiscale Model. Simul.* **2012**, *10*, 986–1022.
- (15) Doll, J. D.; Plattner, N.; Freeman, D. L.; Liu, Y.; Dupuis, P. Rare-event sampling: Occupation-based performance measures for parallel tempering and infinite swapping Monte Carlo methods. *J. Chem. Phys.* **2012**, *137*, 204112.
- (16) Lu, J.; Vanden-Eijnden, E. Infinite swapping replica exchange molecular dynamics leads to a simple simulation patch using mixture potentials. *J. Chem. Phys.* **2013**, *138*, 84105.
- (17) Plattner, N.; Doll, J. D.; Meuwly, M. Overcoming the rare event sampling problem in biological systems with infinite swapping. *J. Chem. Theory Comput.* **2013**, *9*, 4215–4224.
- (18) Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **2002**, *117*, 6911.
- (19) Predescu, C.; Predescu, M.; Ciobanu, C. V. The incomplete beta function law for parallel tempering sampling of classical canonical systems. *J. Chem. Phys.* **2004**, *120*, 4119–4128.
- (20) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. Feedback-optimized parallel tempering Monte Carlo. *J. Stat. Mech.: Theory Exp.* **2006**, *2006*, 3018.
- (21) Sabo, D.; Meuwly, M.; Freeman, D. L.; Doll, J. D. A constant entropy increase model for the selection of parallel tempering ensembles. *J. Chem. Phys.* **2008**, *128*, 174109.
- (22) Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **1995**, *52*, 2893–2906.
- (23) Müller, E. M.; de Meijere, A.; Grubmüller, H. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.* **2002**, *116*, 897.

- (24) Tsallis, C. Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
- (25) Fukuda, I.; Nakamura, H. Deterministic generation of the Boltzmann–Gibbs distribution and the free energy calculation from the Tsallis distribution. *Chem. Phys. Lett.* **2003**, *382*, 367–373.
- (26) Kim, J. G.; Fukunishi, Y.; Nakamura, H. Dynamical origin of enhanced conformational searches of Tsallis statistics sampling. *J. Chem. Phys.* **2004**, *121*, 1626–1635.
- (27) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.
- (28) Piela, L.; Kostrowicki, J.; Scheraga, H. A. On the multiple-minima problem in the conformational analysis of molecules: Deformation of the potential energy hypersurface by the diffusion equation method. *J. Phys. Chem.* **1989**, *93*, 3339–3346.
- (29) Ma, J.; Straub, J. E. Simulated annealing using the classical density distribution. *J. Chem. Phys.* **1994**, *101*, 533.
- (30) Pappu, R. V.; Hart, R. K.; Ponder, J. W. Analysis and application of potential energy smoothing and search methods for global optimization. *J. Phys. Chem. B* **1998**, *102*, 9725–9742.
- (31) Doll, J. D.; Gubernatis, J. E.; Plattner, N.; Meuwly, M.; Dupuis, P.; Wang, H. A spatial averaging approach to rare-event sampling. *J. Chem. Phys.* **2009**, *131*.
- (32) Plattner, N.; Doll, J. D.; Meuwly, M. Spatial averaging for small molecule diffusion in condensed phase environments. *J. Chem. Phys.* **2010**, *133*.
- (33) Brooks, C. L.; Mackerell, A. D.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (34) Wales, D. J. *Energy Landscapes: Applications to Clusters, Biomolecules, and Glasses*; Cambridge University Press: Cambridge, U.K., 2003.
- (35) Eleftheriou, M.; Kim, D.; Doll, J. D.; Freeman, D. L. Information theory and the optimization of Monte Carlo simulations. *Chem. Phys. Lett.* **1997**, *276*, 353–360.
- (36) Efron, B. Bootstrap Methods—Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26.
- (37) Nangia, S.; Jasper, A. W.; Miller, T. F., III; Truhlar, D. G. Army ants algorithm for rare event sampling of delocalized nonadiabatic transitions by trajectory surface hopping and the estimation of sampling errors by the bootstrap method. *J. Chem. Phys.* **2004**, *120*, 3586–3597.
- (38) Nutt, D.; Meuwly, M. Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO. *Biophys. J.* **2006**, *90*, 1191–1201.
- (39) Hu, J.; Ma, A.; Dinner, A. R. Monte Carlo simulations of biomolecules: The MC module in CHARMM. *J. Comput. Chem.* **2006**, *27*, 203–216.
- (40) Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (41) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, *284*, 835–848.
- (42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (43) Ceperley, D. M.; Dewing, M. The penalty method for random walks with uncertain energies. *J. Chem. Phys.* **1999**, *110*, 9812–9820.
- (44) Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.
- (45) Jones, J. E. On the determination of molecular fields. II. From the equation of state of a gas. *Proc. R. Soc. London* **1924**, *106*, 463–477.
- (46) Wales, D. J.; Doye, J. P. K.; Dullweber, A.; Hodges, M. P.; Naumkin, F. Y.; Calvo, F.; Hernández-Rojas, J.; Middleton, T. F. The Cambridge Cluster Database: <http://www-wales.ch.cam.ac.uk/CCD.html> (accessed Aug. 13, 2014).
- (47) Wales, D. J.; Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (48) Schelstraete, S.; Verschelde, H. Finding minimum-energy configurations of Lennard-Jones clusters using an effective potential. *J. Phys. Chem. A* **1997**, *101*, 310–315.
- (49) Doye, J. P. K.; Miller, M.; Wales, D. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* **1999**, *110*, 6896–6906.
- (50) Doye, J. P. K.; Miller, M.; Wales, D. Evolution of the potential energy surface with size for Lennard-Jones clusters. *J. Chem. Phys.* **1999**, *111*, 8417–8428.
- (51) Xiang, Y.; Cheng, L.; Cai, W.; Shao, X. Structural distribution of Lennard-Jones clusters containing 562 to 1000 atoms. *J. Phys. Chem. A* **2004**, *108*, 9516–9520.
- (52) Calvo, F.; Doye, J. P. K.; Wales, D. J. Quantum partition functions from classical distributions: Application to rare-gas clusters. *J. Chem. Phys.* **2001**, *114*, 7312–7329.
- (53) Honeycutt, J. D.; Andersen, H. C. Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. *J. Phys. Chem.* **1987**, *91*, 4950–4963.
- (54) Neirotti, J. P.; Calvo, F.; Freeman, D. L.; Doll, J. D. Phase changes in 38-atom Lennard-Jones clusters. I. A parallel tempering study in the canonical ensemble. *J. Chem. Phys.* **2000**, *112*, 10340–10349.
- (55) Calvo, F.; Neirotti, J. P.; Freeman, D. L.; Doll, J. D. Phase changes in 38-atom Lennard-Jones clusters. II. A parallel tempering study of equilibrium and dynamic properties in the molecular dynamics and microcanonical ensembles. *J. Chem. Phys.* **2000**, *112*, 10350–10357.
- (56) Sharapov, V.; Mandelshtam, V. Solid–solid structural transformations in Lennard-Jones clusters: Accurate simulations versus the harmonic superposition approximation. *J. Phys. Chem. A* **2007**, *111*, 10284–10291.
- (57) Sharapov, V.; Meluzzi, D.; Mandelshtam, V. Low-temperature structural transitions: Circumventing the broken-ergodicity problem. *Phys. Rev. Lett.* **2007**, *98*, 105701.
- (58) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.
- (59) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.
- (60) Wales, D. J. Surveying a complex potential energy landscape: Overcoming broken ergodicity using basin-sampling. *Chem. Phys. Lett.* **2013**, *584*, 1–9.
- (61) Adjanor, G.; Athènes, M.; Calvo, F. Free energy landscape from path-sampling: Application to the structural transition in LJ38. *Eur. Phys. J. B* **2006**, *53*, 47–60.
- (62) Oakley, M. T.; Johnston, R. L.; Wales, D. J. Symmetrization schemes for global optimization of atomic clusters. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3965–3976.
- (63) Tobias, D. J.; Brooks, C. L. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. *J. Phys. Chem.* **1992**, *96*, 3864–3870.
- (64) Apostolakis, J.; Ferrara, P.; Caflisch, A. Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *J. Chem. Phys.* **1999**, *110*, 2099.
- (65) Chekmarev, D. S.; Ishida, T.; Levy, R. M. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models. *J. Phys. Chem. B* **2004**, *108*, 19487–19495.
- (66) Ma, A.; Dinner, A. R. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B* **2005**, *109*, 6769–6779.
- (67) Gfeller, D.; De Los Rios, P.; Caflisch, A.; Rao, F. Complex network analysis of free-energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817–1822.
- (68) Yang, L.; Yi, Q. G. A selective integrated tempering method. *J. Chem. Phys.* **2009**, *131*, 214109.

- (69) Multiple state transition interface sampling of alanine dipeptide in explicit solvent. *J. Chem. Phys.* **2011**, *135*, 145102.
- (70) García-Prieto, F. F.; Fdez Galván, I.; Aguilar, M. A.; Martn, M. E. Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology. *J. Chem. Phys.* **2011**, *135*, 194502.
- (71) Lee, I. H. Free-energy profile along an isomerization pathway: Conformational isomerization in alanine dipeptide. *J. Korean Phys. Soc.* **2013**, *62*, 384–392.
- (72) Morishita, T.; Itoh, S. G.; Okumura, H.; Mikami, M. On-the-fly reconstruction of free-energy profiles using logarithmic mean-force dynamics. *J. Comput. Chem.* **2013**, *34*, 1375–1384.
- (73) Kondo, H. X.; Taiji, M. Enhanced exchange algorithm without detailed balance condition for replica exchange method. *J. Chem. Phys.* **2013**, *138*, 244113.
- (74) Lankau, T.; Yu, C.-H. A constrained reduced-dimensionality search algorithm to follow chemical reactions on potential energy surfaces. *J. Chem. Phys.* **2013**, *138*, 214102.
- (75) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 54103.
- (76) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 20603.
- (77) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615–1621.
- (78) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247–2254.
- (79) Strodel, B.; Wales, D. J. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.* **2008**, *466*, 105–115.
- (80) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, *7*, 95–99.
- (81) Schaefer, M.; Bartels, C.; Leclerc, F.; Karplus, M. Effective atom volumes for implicit solvent models: Comparison between Voronoi volumes and minimum fluctuation volumes. *J. Comput. Chem.* **2001**, *22*, 1857–1879.
- (82) Calimeti, N.; Schaefer, M.; Simonson, T. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins* **2001**, *45*, 144–158.
- (83) Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **2010**, *6*, 35–47.
- (84) Sutto, L.; DÁbramo, M.; Gervasio, F. L. Comparing the efficiency of biased and unbiased molecular dynamics in reconstructing the free energy landscape of Met-Enkephalin. *J. Chem. Theory Comput.* **2010**, *6*, 3640–3646.
- (85) Zhou, T.; Caflisch, A. Free energy guided sampling. *J. Chem. Theory Comput.* **2012**, *8*, 2134–2140.
- (86) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. *J. Chem. Theory Comput.* **2013**, *9*, 1885–1895.
- (87) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (88) Scarsi, M.; Apostolakis, J.; Caflisch, A. Comparison of a GB solvation model with explicit solvent simulations: Potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. *J. Phys. Chem. B* **1998**, *102*, 3637–3641.
- (89) *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2002; Vol. 123.
- (90) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (91) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (92) Jiang, W.; Luo, Y.; Maragliano, L.; Roux, B. Calculation of free energy landscape in multi-dimensions with hamiltonian-exchange umbrella sampling on petascale supercomputer. *J. Chem. Theory Comput.* **2012**, *8*, 4672–4680.