



Published in final edited form as:

*Annu Rep Comput Chem.* 2009 January 1; 5: 49–76. doi:10.1016/S1574-1400(09)00503-9.

## Methods for Monte Carlo simulations of biomacromolecules

Andreas Vitalis<sup>\*,\*\*</sup> and Rohit V. Pappu<sup>\*,\*\*\*</sup>

<sup>\*</sup>Department of Biomedical Engineering, Molecular Biophysics Program, Center for Computational Biology, Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, MO 63130-4899, USA

### Abstract

The state-of-the-art for Monte Carlo (MC) simulations of biomacromolecules is reviewed. Available methodologies for sampling conformational equilibria and associations of biomacromolecules in the canonical ensemble, given a continuum description of the solvent environment, are reviewed.

Detailed sections are provided dealing with the choice of degrees of freedom, the efficiencies of MC algorithms and algorithmic peculiarities, as well as the optimization of simple movesets. The issue of introducing correlations into elementary MC moves, and the applicability of such methods to simulations of biomacromolecules is discussed. A brief discussion of multicanonical methods and an overview of recent simulation work highlighting the potential of MC methods are also provided. It is argued that MC simulations, while underutilized biomacromolecular simulation community, hold promise for simulations of complex systems and phenomena that span multiple length scales, especially when used in conjunction with implicit solvation models or other coarse graining strategies.

### Keywords

Monte Carlo simulations; Polypeptides; Polynucleotides; Concerted rotations; Multicanonical ensemble; Torsional space

## 1. Introduction

Unlike in other fields of computational science, Monte Carlo (MC) sampling is not used very often for simulating biomacromolecules. Historically, this can be explained by a strong bias toward all-atom models of biomacromolecular systems including both macromolecule(s) and solvent. All physically realistic Hamiltonians employ excluded volume and other stiff, non-bonded interactions leading to very rugged energy landscapes and large correlations between many degrees of freedom in these dense systems [1,2]. For such systems, gradient-based techniques such as molecular dynamics (MD [3]) are assumed to be vastly superior, although this notion has been challenged [4]. With the advent of refined implicit solvent models [5-7] and the resultant smoother energy landscapes, MC simulations may gain in appeal for the simulation community. Here, we provide an overview of the state-of-the-art MC methods designed for efficient sampling of biomacromolecules using implicit solvent models in an effective isothermal-isochoric (NVT) ensemble. The use of MC for simulating chains in other ensembles or for simulations of all-atom condensed-phase systems is only touched upon briefly. For more general purposes, the textbook by Frenkel and Smit [8] remains an excellent resource as do other overviews in similar textbooks [9,10].

<sup>\*\*</sup>Correspondence should be addressed to Andreas Vitalis at gilthe.pc@gmail.com or

<sup>\*\*\*</sup>Rohit Pappu, pappu@wustl.edu.

The Markov chain Metropolis scheme [11] is by far the most common MC methodology. The system is randomly perturbed and the proposed move from microstate  $A$  to  $B$  is accepted with probability:

$$p_{A \rightarrow B} = \max [ 1, \exp (-\beta \cdot \Delta E) ] \quad (1)$$

Here,  $\Delta E$  is the energy difference to transition from state  $A$  to  $B$  and  $\beta$  is the reciprocal thermal energy. Metropolis *et al.* [11] showed that such a scheme samples the Boltzmann distribution associated with the given Hamiltonian at the temperature specified by  $\beta$ . For larger systems, such importance sampling is vastly superior to any systematic or random enumeration schemes which scale extremely poorly with the number of degrees of freedom in the system [8].

In principle, the transition from state  $A$  to  $B$  needs to be unbiased and ergodic [12,13]. An illustrative example comes from the example of  $n$ -butane, where the degree of freedom for MC sampling is the single backbone dihedral angle,  $\phi$ . Transitions between microstates are made by randomly choosing values for  $\phi$  from an interval. If, some values of  $\phi$  are chosen with higher probability, then the proposed transitions are sampled from a biased – as opposed to an unbiased – distribution. Conversely, the sampling could be unbiased, but the values for  $\phi$  may be sampled from a small sub-interval of possibilities. In this case, the simulation would suffer from broken ergodicity because the simulation cannot – by fiat – sample all possible conformations for  $n$ -butane. There is a hidden ergodicity issue that pertains to the choices one makes for the degrees of freedom in an MC simulation, a point that is addressed in Section 2.1.

Given a choice for the degrees of freedom, the MC movesets dictate how transitions between different microstates are realized. For biomacromolecules, the movesets have to be diverse. This is important because broken ergodicity can also result from movesets that fail to connect distinct points in conformational space. In any MC simulation the frequencies with which different elementary moves are attempted *and* the parameters associated with the different move types (such as the maximal displacements associated with different moves) are adjustable. If the movesets are unbiased and ergodic, a sufficiently long MC simulation should ultimately converge to the same result. In practice, computational resources are finite and the choices made for adjustable parameters play an important role in determining whether a simulation actually yields a converged result. Metrics to be used to guide choices for these parameters are discussed in Section 2.3. Two important guidelines are as follows. First, as many choices as possible should be made randomly rather than with a pre-determined “schedule”. For example, the choice of degree(s) of freedom to perturb during an elementary move should be random rather than scanning all degrees of freedom systematically. Second, it is always true that every problem has its own optimal parameter set. Hence, simulators should always have the freedom to adjust all open parameters.

There are several other reasons for the scarce use of MC in simulations of biomacromolecules. One is the absence of suitable software. The commercial software, BOSS/MCPRO, is provided by the Jorgensen group [14]. The most common, freely-available simulations packages tailored to the biomacromolecular simulation community, *i.e.*, GROMACS [15], NAMD [16], and TINKER [17], have no MC capability. AMBER [18], which may be the most widely used molecular simulations software, does not provide MC support. However, the CHARMM [19] package now includes an MC module [20]. Some freely available MC programs like MCCCOS Towhee [21] are not specifically tailored to biomacromolecular simulations. Others, like PROFASI [22], currently support only very limited Hamiltonians. As detailed in Section 2.2, the software layout for MD and MC codes is fundamentally different, a reason that contributes to the lack of available programs. We hope that our freely available CAMPARI

software package, which is scheduled for release in the summer of 2009, will provide a useful addition to the small group of suitable programs such as CHARMM, MCPRO, and PROFASI.

Although MC is an underutilized tool in the field of computational molecular biophysics, there are beneficial features that can be exploited, especially in conjunction with implicit solvent models. Specifically, MC has the potential of accessing length scales that are inaccessible to molecular dynamics in complex phenomena like peptide aggregation or conformational sampling of intrinsically disordered proteins.

The rest of this review article is structured as follows:

The first issue we discuss is the choice of degrees of freedom. We briefly review the literature pertaining to justification and implementation of performing molecular simulations in non-Cartesian space, most prominently torsional and rigid-body space. We lay out advantages of such a procedure and comment upon common implementation difficulties, with specific attention to the issue of consistency between the development and application of force field parameters for use in MC simulations.

The second area we review is that of computational efficiency of energy evaluations including special considerations for cutoffs and long-range treatment of electrostatic interactions. We provide a brief overview of the required bookkeeping given a Hamiltonian of a certain complexity. Many of these issues are unique to MC calculations due to the fundamentally different ways in which systems evolve when compared to MD calculations.

Next, we discuss strategies to make maximal use of simple unbiased movesets for conformational sampling of biomacromolecules. We provide an example to illustrate how the inclusion of conformational fluctuations spanning multiple length scales can improve the quality of sampling. We conclude by a discussion of the metrics and guidelines that can be used to optimize a moveset for a given system.

The fourth area we cover is that of introducing correlations into MC movesets. We discuss typical biased movesets employed in MC simulations of biomacromolecules and the corrections and parameter settings needed for the incorporation of biased moves. We address concerns pertaining to computational efficiency and ease of implementation.

The fifth section provides a brief description of multicanonical techniques and their use and applicability in MC simulations. We cursorily touch upon the benefits that MC techniques provide in sampling ensembles other than the canonical ensemble for biomacromolecular systems.

We conclude by providing a few examples illustrating the peculiarities of sampling phase space via MC for non-trivial systems relevant to the biomacromolecular field. We provide an outlook regarding current challenges and the potential strategies that can be developed or adopted to overcome these challenges.

## **2. Conformational sampling of biomacromolecules in the canonical ensemble via Monte Carlo methods**

### **2.1 Choosing degrees of freedom in conjunction with the force field**

One of the benefits of MC algorithms is the ability to naturally deal with constraints, *i.e.*, to set a simulation up in arbitrary sets of degrees of freedom, which may well be different from the degrees of freedom over which the potential energy is evaluated. In MD, such functionality is introduced by holonomic constraints [23], for which a variety of popular algorithms have

been introduced. A long-standing issue concerned the introduction of mass-metric tensor artifacts that might arise if one were to restrict sampling to torsional degrees of freedom alone. These artifacts are undoubtedly present in MD simulations, because momenta conjugate to the constrained degrees of freedom are explicitly set to zero, thereby introducing a spurious alteration of the volume element of the configurational integral by the determinant of a reduced mass-metric tensor. Hence, the efficiency one might gain through the use of longer time steps is lost through the inefficiency associated with calculating mass-metric tensor determinants and its gradients. However, MC calculations rest on the separation between the momentum and configurational integrals and hence the spurious coupling between conformational degrees of freedom and volume elements does not arise because there are no momenta to be zeroed out [24]. A simple test illustrates this point: Consider an  $n$ -alkane; set the potential energy to be zero for all values of the relevant degrees of freedom; randomly sample the torsional degrees of freedom; as long as the torsions are sampled from an unbiased distribution, it should follow that all values for the degrees of freedom have equal likelihoods of being realized. This test reveals that Fixman-style [25] corrections are not needed in MC simulations for chain molecules – an advantage that does not prevail for MD simulations with holonomic constraints.

Integrands of configurational integrals are proportional to  $\exp[-\beta U(\text{DoF})]$ , where  $\beta$  denotes the inverse thermal energy parameter and  $U(\text{DoF})$  refers to the potential energy that varies as the degrees-of-freedom (DoF) assume different values. If  $U(\text{DoF})$  is set to zero, then the multidimensional integral over phase space volumes provides an estimate of the size of the relevant conformational space and all combinations for the degrees-of-freedom should have the same probabilities of being accessed. Any accurate MC algorithm has to reproduce the appropriate unbiased distribution when  $U(\text{DoF})$  is set to zero. This sanity check is very useful since it identifies the presence of bias and the possibility of broken ergodicity in a straightforward manner. For example, if the DoF were Cartesian coordinates, and one were using periodic boundary conditions, the degrees-of-freedom should have a uniform distribution in all three dimensions. Torsional degrees of freedom behave similarly. This makes it possible to sample dihedral angles in MC simulations from a uniform random distribution. The same is not true for degrees-of-freedom where the configurational space volume element depends on the values assumed by the degrees-of-freedom. Examples of such degrees-of-freedom are Euclidean distances and Euler angles [10]. In these cases, Jacobian corrections need to be included in either the acceptance or picking probabilities.

The vast majority of MC simulations of biomacromolecules employ torsional space sampling which is sometimes augmented by sampling of angular degrees of freedom [26] or even the Cartesian coordinates directly [27]. In the presence of stiff harmonic restraints, step sizes for the latter will typically be vanishingly small. Inclusion of such moves is not a matter of choice. Instead, it is determined by the force field used for the calculations. Every molecular mechanics force field, whether it uses an implicit or explicit representation of the solvent, undergoes a calibration process. Rigorously speaking, the applicability of said force field is only guaranteed within the accuracy of calibration if the sampled degrees of freedom are strictly identical between application and calibration. For instance, introducing holonomic constraints on bond lengths and angles in MD simulations using a force field parameterized in the absence of such constraints is incorrect. Rotational barriers in peptides are known to depend strongly on the flexibility of the stiffer modes [28], including the peptide bond itself [29]. While it is argued in general that equilibrium properties are not affected by typical constraints in a statistically significant manner, even the impact on barriers and hence kinetics might be enough to question their introduction into a force field parameterized in their absence.

The consequences for MC simulations of biomacromolecules are obvious: The choice is to either employ a force field designed specifically in the presence of such constraints (for example [30]), or – using a diverse enough moveset – to make sure to sample the appropriate

coordinate space directly. The consequences of ignoring this concern are shown in Figure 1. Here, we compare the temperature-dependent, reversible folding / unfolding of an  $\alpha$ -helical peptide (Ace-A<sub>5</sub>(AAARA)<sub>3</sub>A-Nme) for the ABSINTH implicit solvation model and Lennard-Jones parameters [30] coupled to OPLS-AA/L partial charge parameters [31]. We compare results from MC in dihedral space (all constraints present) to results obtained using Langevin dynamics (LD [32]) in Cartesian space. The latter used no constraints whatsoever and bond length and angle parameters were ported from the OPLS-AA/L force field [31]. The MC methodology is identical to what has been published previously [30] while the LD methodology is described in the caption to Figure 1. As can be seen, there is a substantial shift in the temperature-midpoint of the melting transition of the  $\alpha$ -helix. The additional flexibility decreases helix stability leading to a lower melting temperature. This result is independent of the metric that we used to quantify the helix-coil transition. It should be noted that some hysteresis error remains for the LD calculation, even though the CPU time needed was larger by a factor of 3-5 than the MC simulation. This efficiency benefit provided by MC sampling is slightly larger than, but generally consistent with what has been reported in the literature for systems of similar size [26].

In this particular example, constraints strongly affect equilibrium properties as quantified in terms of the response of the system to changes in temperature. This example shows that it is not easy to decouple the effects of constraints on enthalpic barriers from those on minima on the free energy landscape. Caution is therefore required when employing typical biomacromolecular force fields such as AMBER [33], CHARMM [34,35], OPLS [31], or GROMOS [36] in MC simulations in dihedral angle space alone. It is important to ensure that the parameters of a forcefields have been calibrated using MC as the sampling engine prior to using these parameters in an MC simulation. While this is not the case for simulations of simple systems such as Lennard-Jones fluids or even associative liquids made of rigid molecules, it will certainly be an important consideration for highly flexible polymeric systems.

## 2.2 Bookkeeping and efficiency in computational algorithms for MC simulations

From a software point of view, it is desirable to have a well-structured hierarchical description for the different biomacromolecules in the simulation system. Such a hierarchy should provide data structures and access functions on the atomic, residue, molecule, and system levels. This allows routines for the evaluation of energy terms to be set up at the level of residue pairs. Experience [14] suggests that this setup is advantageous since it provides a route to easily and intuitively implement the computational algorithms sketched below.

The requirements for data structures and bookkeeping of simulation variables differ fundamentally between MD and MC simulations. In MD programs, forces and energies involving all degrees of freedom are calculated at every step. The system is subsequently evolved and this process is then repeated for the duration of the simulation. The two core assumptions, which are taken advantage of, are i) all force and energy evaluations are global, *i.e.*, are performed for the whole system, and ii) the system evolves in incremental steps such that there is high correlation between the forces and energies from step  $n$  to step  $n+1$ . These assumptions give rise to many algorithmic strategies that are used to speed up MD calculations. These strategies include the use of twin-range cutoffs, neighbor lists which are updated infrequently, and the particle-mesh Ewald (PME [37]) method for treating corrections to long-range interactions.

Unfortunately, neither assumption is true in MC calculations. Instead of requiring global evaluations of forces, evaluations of energy differences  $\Delta E$  are needed between pairs of microstates that are not necessarily close in phase space, but which might only differ in the values for a subset of degrees of freedom. This implies that a majority of the energy terms remain fixed and they need not be considered when computing  $\Delta E$ . For the sake of efficiency,

this necessitates a strategy for incremental energy updates that is customized for each individual MC move type, which is part of the available moveset. As an example, consider a solution of several small molecules as sketched in Panel A of Figure 2. If we assume a pairwise additive Hamiltonian and no cutoffs, the terms of the Hamiltonian which require re-computation represent only a small subset of the terms needed to compute the whole system energy. The computational expense for energy evaluations in this case scales as  $O(N)$  with the number of interaction sites  $N$ , if the number of molecules is large relative to the number of interaction sites per molecule. For biomacromolecules and their associated movesets, it is important to analyze the efficiency of different elementary move types in these terms. For example, consider the perturbation of an individual backbone dihedral angle in a single biomolecule (Panel B of Figure 2). Depending on the location of the residue, the worst complexity for incremental energy evaluations for this type of move is  $\sim N^2/2$ . For larger molecules, such moves become increasingly less efficient due to the inherent  $O(N^2)$  scaling. Conversely, the complexity for a perturbation of a single sidechain degree of freedom is only of order  $O(N)$  (analogous to Panel A of Figure 2). Even with cutoffs, the scaling in computational complexity for energy evaluations between different elementary move types will be different. These considerations provide partial motivation for the development and inclusion of truly local MC moves for which scaling ultimately is  $O(N)$  (see Section 2.4.1).

Aside from the programming complexity of writing separate wrapper routines for energy updates, the following considerations are peculiar to MC calculations: Not every energy function that is used in simulations of biomacromolecules is suitable for decomposition into static and changing terms. Obviously, a strictly pairwise-additive Hamiltonian is well suited for this type of decomposition. Conversely, non-pairwise-additive Hamiltonians give rise to the complexity that an interaction between two sites is in fact changed as the result of the movement of a third site. This is the case for many implicit solvent models such as the Poisson-Boltzmann (PB [5]) and generalized Born (GB [6]) models, later generations of the EEF1 model [38] and the ABSINTH model [30]. For these models, the effective multibody interactions have to be subjected to a cutoff or similar simplifying assumption to allow efficient use in MC simulations.

In the ABSINTH implicit solvent model, the range of multi-body interactions is limited by the size of each atom's (implicit) solvation shell. Since the coupling parameter for these interactions is the solvent-accessible volume, we can keep track of which interactions to recompute in addition to those involving the moving parts during that elementary MC move. We do so by identifying and marking all atoms whose solvent-accessible volumes have changed as a result of the move. All interactions involving at least one of those atoms are then recomputed as well. The strictly local nature of the three-body interactions makes this treatment exact for the ABSINTH model. The scheme also integrates naturally with pairwise cutoffs. Conversely, alternative strategies were proposed for the GB model introducing cutoffs directly into the coupling terms based on their magnitude [39,40]. To correct such an inaccurate treatment in MC simulations, it is important to point to a general strategy cast differently by Gelb [41] and Hetenyi *et al.* [42], which allows the bulk of Metropolis sampling to occur from either a simplified potential or a fast-varying subset of the potential. These short simulation stretches are periodically accepted or rejected using an "outer" Markov chain in order to ensure that the desired distribution happens from the full Hamiltonian.

A similar strategy can potentially be used to calculate electrostatic interactions without cutoffs in periodic boundary conditions using the popular PME method [43]. In MC, the same problem arises as outlined in the previous paragraph due to the non-decomposability of the reciprocal space sum. Again, a system-wide evaluation of the full potential energy including the reciprocal part would only occur in regular or random intervals while the bulk of the sampling would be performed from the Hamiltonian given by the real-space sum alone. An analogous

implementation for the standard Ewald method has been presented [44]. Conversely, direct use of the Ewald sum [45] or approximations to it [46-48] which are pairwise decomposable and hence suitable for MC simulations have generally proven to be too inefficient for most modern applications [49]. Additionally, it should be pointed out that Ewald sums – independent of implementation – are incompatible with implicit solvent models that model a spatially varying dielectric with anything more than trivial functional dependencies [45].

The issue of electrostatics brings up the issue of cutoffs in general. It is beyond the scope of this review to discuss the artifacts introduced by using cutoffs on the electrostatic, excluded volume, and dispersion interactions. It is sufficient to state that for simulations of polar condensed phases or of biomacromolecules in an implicit solvent model such artifacts are hard to detect in the canonical ensemble for system volumes of sufficient size. This stands in contrast to condensed phase systems with mobile charges (see [50] for an illustration). Regardless, for the MC simulations to remain scalable to larger systems, it is inevitable that some form of simplification of the infinite-range Hamiltonian be considered. A necessary step for distance-dependent potentials is the generation of neighbor lists, *i.e.*, the efficient determination of spatial proximity relationships. Here, MC simulations can take advantage of two common strategies:

- a. Grid-based methods which operate in  $O(N)$  time predominantly known as linked-cell lists [51], which are usually implemented at the level of atoms.
- b. Hierarchical methods which take advantage of prior knowledge about molecular topology and which operate in  $O(M^2)$  time where  $M = N/n$  and  $n$  is the (mean) number of atoms per repeating unit in a polymer, *e.g.*, a single protein residue.

With either method, for an energy calculation associated with an elementary move in a large system, a vast majority of the interatomic distances involving the moving parts are never computed. In contrast, Verlet lists [52] are only useful in MC calculations of dense fluids due to the local nature of elementary moves in this particular case.

In summary, strategies developed for MD simulations do not usually apply in MC simulations. Neighbor lists and energies need to be incrementally updated for the trial move and kept or restored to their original values in case of an accepted or rejected move, respectively. Strong emphasis should be placed on computational efficiency. First and foremost only terms which need to be computed should be computed. As with all software, the complexity of algorithms consuming the bulk of CPU time should be carefully analyzed and minimized. Furthermore, strategies using a simplification of the potential and subsequent correction methods or  $O(N)$  movesets represent promising avenues for future development. Small enough systems, however, can be dealt with comfortably using even simple movesets. This is the focal point of Section 2.3.

### 2.3 Optimizing simple MC movesets for biomacromolecular simulations

With the prominent degrees of freedom being dihedral angles (see Section 2.1), a set of straightforward movesets emerges. Polypeptide backbone and sidechain torsions including the peptide bond, but excluding rigid rings or planar systems in the sidechains are sampled from a uniform prior distribution. The number of degrees of freedom for each peptide residue therefore ranges from three (Gly) to seven (Lys). For polynucleotides, dihedral angles along the phosphate-sugar backbone are sampled along with rotations around the nucleoside bond as well as any existing carbon-oxygen bonds involving free OH-groups [53]. These degrees of freedom can be sampled using simple perturbations of individual or multiple dihedral angles. If such a dihedral angle is part of the polymer backbone, one end of the chain pivots around this joint. Hence, such moves are commonly referred to as pivot moves [54]. In general, moves involving simultaneous perturbations of several degrees of freedom decay exponentially in

efficiency with the number of degrees of freedom for a realistic Hamiltonian. This is due to the linearly growing chance of encountering steric clashes, which goes into the energy difference governing the exponential Metropolis criterion (see Equation 1). Nonetheless, they can be useful as they randomly introduce correlation. The latter may be necessary to convert between conformational states for highly coupled cases like sidechain rotamers [55].

Non-aromatic rings require special treatment. Significant conformational flexibility is retained and a rigid description is inappropriate. The different pucker states of the sugar moiety or the proline sidechain are however characterized by a relatively discrete ensemble as is evidenced by analysis of high-resolution crystal structures [56-58]. Simple approaches therefore can be designed and implemented by creating a discrete set of states with an associated energy function meant to reproduce the proper distribution in the context of a suitable background Hamiltonian. Alternatively, specific approaches to find solutions to the ring closure problem may be implemented [59]. Such algorithms are discussed in a different context in Section 2.4.1.

As is well-known, the dynamics of polymers in a dense environment under a broad range of conditions become very slow, often glassy [60-63]. A canonical example is that of a single, long polymer in a poor solvent, specifically a mean-field solvent environment in which self-interactions are preferred over chain-solvent interactions [64]. This problem provides a prototype of the complexity encountered in protein folding problems with implicit solvent models, and hence is of significant relevance to the biomacromolecular field [65]. Aside from all more complex movesets inspired by such sampling problems (see Section 2.4), what can be done to improve the MC methodology even using simple moves alone? And how does one establish metrics to track sampling efficiency?

We demonstrate the efficiency of a straightforward advancement in simple MC movesets using the peptide Ace-Nle<sub>30</sub>-Nme under “deep quench” conditions. Here, Nle indicates the norleucine residue which is isosteric with lysine. We monitor the collapse of this peptide from a fully extended state using a degenerate, poor solvent Hamiltonian. Such a Hamiltonian is provided by only employing Lennard-Jones interactions according to parameters published previously [30]. A universal parameter for MC moves is the step size, which is normally sampled from a uniform distribution over a finite interval. Figure 3 compares three different sampling approaches to pivot moves: i) All dihedral angles are completely randomized each time they are sampled (maximum step size), ii) all dihedral angles are perturbed in stepwise fashion, and iii) both methods are mixed. As can be adjudicated from the relaxation behavior of the system, the efficiency of approaches i) and ii) does seem to track with the density of the system. Full randomizations perform well in the low density limit, but poorly in the collapse regime. The opposite is true for stepwise perturbations. The important point is that sampling of multiple length scales provides a rigorous, synergistic benefit. The strategy of splitting the moveset into multiple variants of the same basic move type introduces more parameters for the simulator to set. Unfortunately, it is usually not possible to optimize movesets for every problem studied, which implies that intuition and rules of thumb combined with preliminary simulations will inevitably remain prevalent in setting up MC simulations. However, if a more rigorous parameter optimization is needed, we need metrics that can be used to report on the efficiency of sampling.

Aside from relaxation measures such as those shown in Figure 3, metrics of sampling inevitably relate to the rate of “conformational diffusion” or “conformational drift”. The first and foremost test in this regard should always be reproducibility by running identical replicas of the same simulation with different starting conditions. Standard deviations of ensemble averages from a sufficient number of independent individual runs yield standard errors, a procedure similar to block averaging, but avoiding all potential correlations between blocks. The magnitude of those standard errors is a high-level, but reliable test to guide the optimization of MC movesets.



Some systems show reversible order-disorder transitions as a function of control parameters such as temperature. Typical examples of such systems / problems are helix-coil transitions in polypeptides, folding-unfolding transitions of globular proteins, melting transitions of small loop-forming RNA systems, and globule-to-coil transitions in flexible polymers. If disorder-to-order transitions in such systems are reversible, then simulated values of order parameters that track these transitions should yield similar values in the forward and reverse directions. A useful test of a converged Monte Carlo simulation protocol is to test for reversibility. If there is hysteresis in the simulated order-disorder transitions, then the simulation has clearly not converged. This hysteresis check is directly adopted from experimental work on two-state systems and represents an excellent test within the reduced-dimensional space the two-state analysis is performed in (see Figure 1). A measure that is similar in spirit is the generalized ergodic measure developed by Straub and Thirumalai [66] which is specifically designed for MD.

Variance- and covariance-based measures such as autocorrelation “times” of instantaneous quantities are useful guides, but are less applicable to MC simulations in particular. First, the absence of significant variance is probably an indicator of a lack of efficient sampling. Second, MC simulations often take discontinuous paths through phase space, which produces substantial stochasticity in such analyses, especially vis-à-vis MD data. Nonetheless, with some prior knowledge of the energy landscape such measures can be used [67]. The bulk of recent work in this area remains dominated by finding efficient ways to cluster simulation data using either principal-component analysis (PCA) [68], direct clustering techniques [69], or other reduced-dimensional quantities [70,71]. It is then assumed that comprehensive sampling in those spaces is possible and that better coverage and more frequent transitions between basins correspond to improved sampling. Such analyses are typically independent of the sampling engine and therefore suitable for usage in optimizing MC movesets.

Finally, for simulations involving multiple molecules, the rigid-body degrees of freedom have to be sampled. Standard approaches in the spirit of the original Metropolis method are feasible for translational displacement. Rotations of asymmetric particles are conveniently handled by quaternions [72]. One of the cutting-edge applications of MC simulations in the biomacromolecular field is the simulation of peptide aggregation at typical *in vitro*, *i.e.*, often very low concentrations. Assuming an implicit solvent representation, the density of such a system is small whereas the volume is large. In MD, this poses the problem of diffusion-limited kinetics, a problem which in certain rare cases may be overcome by adaptive time step methods [73-75]. The concept of sampling multiple length scales simultaneously applies here in an even more straightforward manner. It is well worth the additional parameters to introduce moves which fully randomize the translational and rotational degrees of freedom of a given molecule. A recent application demonstrates the usefulness of such an approach for the reliable sampling of polyglutamine dimerization at a very low concentration [76].

Of course, a simple moveset will eventually become inefficient with increasing complexity (density and size) of the problem. Hence, diverse attempts have been made to design new and better movesets for MC simulations. Those are discussed next.

## 2.4 Introducing correlation into MC movesets

As touched upon above, the strategy of introducing correlations into Monte Carlo simulations by simultaneously changing multiple degrees of freedom is a losing proposition due to increased combinatorial complexity. This is because pivots about multiple torsions lead to rejected moves, especially when the chains are in dense phases. However, correlations are necessary because the simple moves have the complication that the conversions between distinct dense phase configurations are not readily sampled with simple pivot moves. Hence, much effort has been devoted to design movesets that are inherently biased toward inducing a

concerted change in the system involving several degrees of freedom at a time. The need for correlation becomes apparent if one considers rugged energy landscapes. More often than not, the paths connecting two adjacent minima will involve a collective degree of freedom, *i.e.*, a concerted change in some or all of the relevant elementary degrees of freedom. Such a scenario is sketched in Panel A of Figure 4 for the case of two elementary degrees of freedom. Even if the path is of finite width, elementary moves perturbing only one of the degrees of freedom at a time would be largely ineffective in connecting the two minima. Conversely, an MC move biased toward steering the system along this path would have high acceptance rates despite the need for correcting biases that are introduced by the sampling (see below).

**2.4.1 Concerted rotation (loop-closure) algorithms**—A common situation that requires the introduction of correlation into MC movesets is the case in which the ends of the macromolecule itself or of stretches within it are constrained. This is the case for any ring system intended for MC sampling. These include the five-membered rings of sugars and proline, chemically cross-linked macromolecules such as proteins with disulfide linkages, and of circular peptides or DNA. However, a much broader range of applications emerges from simply considering a consecutive stretch of residues within a macromolecule. The enclosed stretch or loop is considered on its own. The conformation of the residues in that stretch is re-sampled while the ends remain in place. These so-called concerted rotation or loop closure algorithms are attractive for three reasons. First, they are truly local and  $O(N)$  and can hence be computationally efficient for a large systems (see Section 2.2). Second, they introduce correlations between multiple degrees of freedom. Third, they lead to local perturbations expected to sample the system much more efficiently than pivot moves at high density, *e.g.*, in the interior of a folded protein.

On a lattice, so-called crankshaft moves are trivial implementations of concerted rotations [77]. They have been generalized to the off-lattice case [78] for a simplified protein model. For concerted rotation algorithms that allow conformational changes in the entire stretch, a discrete space of solutions arises when the number of constraints is exactly matched to the available degrees of freedom. The much-cited work by Gō and Scheraga [79] formulates the loop-closure problem as a set of algebraic equations for six unknowns reducible effectively to a single equation in a single unknown. The latter is solvable by a systematic search process and eventually yields one or more discrete solutions for all six unknowns. This approach has been recast, modified and extended multiple times to design algorithms specifically suited to allow local MC moves or exhaustive loop sampling for biopolymers [80-82].

In the context of local MC moves, which are of principal relevance here, we can break the procedure down into several stages:

- a. A biased or un-biased pre-sampling or pre-rotation step perturbing a part of the chain.
- b. The chain closure algorithm solving the constraint problem for six additional degrees of freedom to close the chain exactly.
- c. Computation of Jacobian weights for the entire composite move.

Step a) is usually included in order to ensure that the resultant conformation is significantly different from the previous one. If step a) were to be omitted, the identity transformation, *i.e.*, the solution represented by the conformation the chain is in originally, would always be obtained as a possible solution which is disadvantageous. The number of degrees of freedom contained in a) is a free parameter. However, random rotations around multiple dihedrals will easily generate conformations with no solution under the assumption of constant bond lengths. Hence, perturbations around the pre-rotation segment are typically either small and / or restricted to very few if not a single dihedral angle [83-85], or biased toward keeping the end of the (longer) pre-rotation segment in roughly the same location [81]. This latter idea was

originally introduced by Favrin *et al.* [86] as an approximate chain-closure technique, a method which on its own suffers from the quasi-local nature of the resultant moves.

Step b) is often the time-consuming one due to the need for performing a systematic search of the solution space in at least one variable. Most implementations suggest a systematic scan of solution space for both forward and reverse moves which is necessary to properly employ Jacobian-based weighting. Two related advances were proposed: Mezei [85] suggested limiting the search space to conformationally close solutions using a “reverse proximity criterion”. It was shown that performance is superior when compared to the more complete version of Hoffmann and Knapp [83]. Similarly, Ulmschneider *et al.* [81] use a restricted search space with their concerted rotation variant sampling three dihedrals and bond angles each to arrive at a single solution on average. In both cases, the Jacobian needs to be computed only for a single forward and reverse transformation rather than for multiple solutions. This step c) was ignored in early work and identified and introduced by Dodd *et al.* [82]

An alternative route to implement local MC moves is provided by the literature on (inverse) kinematics, such as on control systems for robotic arms composed of flexible joints [27,87]. Here, the problem is transformed to either a set of linear equations [27] or finding the roots of a high-order polynomial [87] at comparable computational expense. One of the benefits of such an approach is the ability to introduce arbitrary stiff segments into the loop, *i.e.*, the degrees of freedom used for chain closure do not have to be consecutive. Conversely, library-based strategies such as that introduced by Kolodny *et al.* [88] are not suitable for MC simulations due to the non-ergodicity of the moveset.

In summary, the recommended implementation for concerted rotation moves in MC simulations uses:

- i. A pre-rotation segment of arbitrary length with controllable bias toward keeping the perturbation small
- ii. An efficient solution of the closure problem for an arbitrary set of non-consecutive degrees of freedom
- iii. Implementations tailored to all biopolymers including polypeptides, polynucleotides, and potentially even lipids as well as polysaccharides

This of course leaves a large set of parameters to optimize for most problems. Only if the simulator has control of all these parameters, can the literature be used to guide those choices. It is beyond the scope of this review to exhaustively test and compare various implementations for MC concerted rotation moves. Naturally, an improvement in the quality of sampling is shown in the original publications for nearly all proposed methods. An interesting data point comes from Ulmschneider *et al.* [26] who compare MC to MD sampling for a set of small proteins or peptides capable of reversible folding. In a GB/SA implicit solvent model with the OPLS-AA/L force field, they find that MC sampling, which consists of concerted rotation and simple sidechain moves, is superior by a factor of 2.0-2.5 using folding times as the dominant metric. This compares favorably with our own experience as detailed in Section 2.1 (Figure 1).

**2.4.2 Cluster move algorithms**—Traditionally, density has been the most crucial limiting factor in making MC techniques useful. At typical liquid densities, acceptance rates even for single-particle moves drop precipitously. Small molecule diffusion is hindered, and the simulation of binary mixtures is near-impossible at those densities, and – more importantly – vastly inferior to MD. The reason is of course the extremely high coupling between all degrees of freedom in the system, for which no efficient MC move sets can be designed (see Figure 4).

At slightly lower densities, however, significant advances have been made to introduce correlation into rigid-body moves. The naïve approach is to randomly select two or more molecules and to perturb their rigid-body degrees of freedom in concerted fashion, *i.e.*, to translate them by the same vector and / or to rotate them around a common point. Such a moveset is perfectly ergodic, unbiased, and potentially able to capture all positive correlation between molecules. Unfortunately, it is highly inefficient and explodes combinatorially with system size. Even for a system of 100 molecules, the chances of picking every possible “cluster” of size four are vanishingly small (there are more than  $10^8$  such unique clusters).

In response, two dominant algorithms were developed initially for on-lattice spin systems [89,90]. Both revolve around determining an effective cluster of spins iteratively based on pairwise energies to construct a move type capable of overcoming the correlation problem in these systems, which were typically observed as and referred to as “critical-slowdown phenomena”. In constructing the clusters, energetic coupling information is therefore used directly rather than inferring it from spatial coupling. The resultant algorithms can be cast in a rejection-free manner for certain Hamiltonians and differ fundamentally from the Metropolis method. The first algorithm is due to Swendsen and Wang [89], the second due to Wolff [90]. The literature dealing with methodological advances for both of these methods has been reviewed recently [91]. Off-lattice variants were iteratively refined [92-95] leading to highly efficient movesets for evolving systems of molecules at low enough density, where “low enough” is typically governed by the range and strength of interaction potentials (see for example [96]). The major drawback of all of these methods is the assumption of a pairwise additive potential. As outlined in Section 2.2, this is not the case for most modern implicit solvation models which rely on effective many-body interactions. Further development is needed to design an efficient strategy addressing or circumventing this issue. The reader should be reminded however, that – as an additional complication – none of the above algorithms is known to work for a condensed phase system of strongly interacting particles with long-range interactions. An alternative formalism is proposed by Maggs [97] which might work better in such circumstances.

What is the utility of this body of literature in the context of all-atom simulations of biomacromolecules? Two cutting-edge applications come to mind: First, simulating the aggregation or transient association of peptides and / or proteins with the single-particle Metropolis method is obviously hindered as soon as strongly interacting molecules are present and associate. Due to the low net density, algorithms as presented above represent elegant ways to circumvent those kinetic traps [94]. Second, we anticipate a rapid growth in mixed explicit-implicit solvent models. A simple case is the explicit treatment of electrolyte ions or other cosolutes coupled to a continuum description of the water alone [30]. Often, the cosolutes may represent a dense enough matrix to slow down diffusion of macromolecules even if cosolute molecules do not specifically bind to the latter. Sampling in such mixed-size solutions may be substantially enhanced using appropriate cluster algorithms [96].

**2.4.3 Gradient-biased Monte Carlo techniques**—One of the oldest ideas in MC simulations is to improve their efficiency by using information about the potential energy gradient. From the outset, this poses two challenges: i) gradients need to be computed eliminating one of the efficiency benefits of MC over MD, and, ii) for rugged energy landscapes, gradients have only local predictive power; *i.e.*, they do not yield guidance for crossing barriers even though the step-size is not principally constrained as it is in MD. Dating back to the works of Pangali *et al.* [98] and Rosky *et al.* [99], single-particle forces have been used to guide the displacement of particles in dense systems. While a considerable speed-up is typically observed relative to the unbiased Metropolis scheme, it remains unclear whether such a method is ultimately superior to dynamics techniques which ensure explicitly that the canonical ensemble is sampled, *i.e.*, stochastic (Langevin) dynamics. Brownian dynamics

(BD), which is stochastic dynamics in the overdamped limit, can just as well be understood as force-biased (dynamic) MC employing collective moves only [100,101].

Not surprisingly, a large body work has emerged trying to link the methodologies while taking full advantage of the correlation introduced by using gradients. Such correlation is maximal for Newtonian MD due to the absence of random noise, which would cause both friction and momentum correlation loss (governed by the fluctuation-dissipation theorem). Originally, Duane *et al.* [102] suggested augmenting MD with Metropolis MC moves to accept only configurations consistent with the canonical ensemble. The requirement in the MD portion is that the integrator be time-reversible and symplectic. This mixing of NVE MD sampling with an outer Markov chain has enabled taking larger time steps in the MD portion compared to straight NVE or NVT MD. The efficiency, however, has been criticized due to rapidly decaying acceptance rates essentially caused by integrator error [103]. Hence, multiple improvements have been suggested [104-108] which extend beyond the scope of this review. An interesting question can be raised: why is NVT MC not simply alternated with dynamics methods ensuring sampling from the same, *i.e.*, canonical distribution? Clearly, Newtonian MD with the widely popular weak-coupling method [109] does not ensure sampling from the proper distribution, but others including Langevin dynamics do. Assuming short momentum autocorrelation times in the presence of significant friction, any error introduced by resetting velocities periodically should be small to negligible. One concern is theoretical in nature and arises from the fact that such an approach cannot be easily cast as a single Markov chain. Another issue might simply be a technical point and relate to considerations outlined in Section 2.2.

We conclude that gradient-based techniques, in particular hybrid MD/MC protocols are of fundamental importance to the biomacromolecular simulation field due to their universality. They are universal in that they are independent of the details of the system as long as the potential energy function is differentiable. While the implementation challenge of deriving and calculating analytical gradients is non-trivial for certain Hamiltonians, such methods present the most intuitive and straightforward route to introduce correlation into the evolution of the system. We therefore recommend that a gradient-based hybrid method which rigorously samples the canonical ensemble be added as an elementary move to available MC software. The aforementioned universality stands in contrast to techniques designed specifically for lattice systems or bead-spring polymer models. Some of those latter techniques and their potential as tools in biomacromolecular MC simulations are discussed next.

**2.4.4 Other advanced techniques and their applicability to biomacromolecular systems**—The polymer literature yields a variety of specialized move types in particular for lattice homopolymers [110]. Sampling methods like the slithering snake and reptation algorithms (see [111] and references therein) or the original configurational-bias / chain growth algorithms [112,113] were specifically designed for lattice representations. Despite their extension to continuum systems and subsequent improvements [114-122], successful applications to biomacromolecules at an all-atom representation have not been reported. Certainly, the complexity of actual biological heteropolymers eliminates typical assumptions about molecular topology and geometry which are taken advantage of in these cases.

However, it also needs to be clarified that not all sampling problems can be solved by introducing correlation to the moveset. To illustrate this point, consider Panel B of Figure 4. Instead of an almost barrier-free path connecting the two minima, it is just as well possible for the barrier to be insurmountable. In this case, stepwise MC and MD methodologies are stunted. This is where multicanonical techniques come into play by utilizing data obtained from systems under different conditions, usually different temperatures. A very brief overview of such techniques is given in Section 2.5. Of course, there are multitudinous techniques which drop the requirement to stringently sample from the correct ensemble and can be more easily

classified as energy landscape exploration tools [123]. Often used in conjunction with library-based approaches, such methods can potentially be extremely useful in guiding the development of new MC movesets, which specifically capture the intricate correlations needed for efficient sampling using a fundamentally random approach.

## 2.5 Beyond the canonical ensemble

The common multicanonical techniques such as replica-exchange or simulated tempering have been described and reviewed extensively in different contexts [124]. They interface naturally with MC simulations as they are cast as (biased or unbiased) random walks in terms of a control parameter – usually temperature. They work by exchanging information between the different conditions thereby allowing increased barrier crossing and quicker convergence of sampling at all conditions of interest.

In addition, a variety of techniques with a narrower focus on enhancing the sampling at a single condition have been developed. There is a variety of techniques employing higher temperature data to work toward that goal, *i.e.*, to reduce the apparent non-ergodicity at the sampling temperature of interest [125-130]. Essentially these techniques can be thought of as forming a continuum as they ultimately rely on similar ideas and/or are based directly on one another. For systems that are of low complexity, the so-called flat-histogram methods [131,132] present another alternative to solve the issue of apparent broken ergodicity. Here, a random walk in energy space is constructed to determine the density of states directly which then yields thermodynamic quantities. These methods still seem to be restricted to simplified systems as a very recent application to a lattice protein folding problem demonstrates implicitly [133]. Similarly, the promise of an extension of the method to include a density bias was demonstrated on a discretized protein model [134].

In summary, a wide variety of tools are available to solve canonical sampling problems by using information from different generalized ensembles. Such techniques are truly independent of the underlying MC methodology, whose review in the context of biomacromolecules forms the bulk of this review article. In the next two sections we conclude by presenting a few recent highlights demonstrating the applicability and usefulness of MC sampling to problems of biophysical and/or physicochemical interest.

## 3. Highlights of MC simulations of biomacromolecules and outlook

Undoubtedly, this paragraph needs to be prefaced by the disclaimer that the MC simulation work cited below is only a sample, which is by no means exhaustive. Shakhovich's group has investigated the folding of several biomacromolecules of interest by coupling MC sampling to a simplified Hamiltonian biased toward the native state, *i.e.*, a Go model [135-137]. Such studies have been quite feasible due to the better compatibility of MC methods with simple potential energy functions. Another example employing statistical potentials comes from Shental-Bechor *et al.* [138].

The work of Irbäck deserves special mention where the application of MC methodology to biomacromolecular systems is considered. Irbäck and collaborators recently applied a simple, efficient, knowledge-based implicit solvent model to a variety of biologically relevant problems [139], ranging from the aggregation of amyloidogenic peptides [140] to the mechanical unfolding of proteins [141]. These are two highlights within a larger body of work [139] that relies exclusively on MC sampling. They demonstrate the potential inherent in combining MC with implicit solvent models.

Similarly, Ulmschneider *et al.* showed that proper MC sampling can be more efficient than MD for the folding of small peptides [26]. An impressive demonstration of the capability of

MC is a recent study on the folding of a transmembrane helix in an implicit membrane environment [142,143]. Vitalis *et al.* [76] demonstrated that MC simulations can indeed bridge length scales (and hence timescales) inaccessible to conventional dynamics techniques. They simulated the dimer formation of two intrinsically disordered polypeptides and obtained converged associativity data at effective concentrations as low as 100  $\mu\text{M}$ . A dynamics-based approach would have been infeasible in this case due to the limitation of molecules having to diffuse hundreds of angstroms.

De Mori *et al.* have taken a different approach to take advantage of MC simulations. They used a coarse-grained Hamiltonian to pre-sample phase space in an approximate manner. This is followed by MD simulations starting from representative structures from the most dominantly populated clusters within the MC ensemble. Such a hierarchical strategy was employed to study the folding of a small protein [144] and the oligomer formation of short, amyloidogenic peptides [145].

We wish to reiterate the theme of combining implicit representations of the solvent environment with all-atom models of the biomacromolecules taking advantage of the sampling benefits of MC. Clearly, the boundaries of computer simulation can be pushed to limits which are not easily reached given finite resources. The time is right for the simulation community to participate in the application and development of MC methodology for biomacromolecular systems.

## Acknowledgments

This work was supported by grants MCB 0718924 from the National Science Foundation and RO1 NS056114 from the National Institutes of Health to RVP. We are grateful to Xiaoling Wang, Adam Steffen, Nicholas Lyle, Hoang Tran, Tim Williamson, Emma Morrison, and Albert Mao for helpful discussions and data generated using the CAMPARI software package (or earlier versions of this package) that helped guide our thinking regarding Monte Carlo simulations. RVP is grateful to Nathan Baker, Alan Grossfield, and Gregory Chirikjian for many helpful discussions over the years.

## 4. Bibliography

1. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Quarterly Reviews of Biophysics* 2002;35:111–167. [PubMed: 12197302]
2. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory part II: Quantitative aspects. *Quarterly Reviews of Biophysics* 2002;35:205–286. [PubMed: 12599750]
3. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature* 1977;267:585–590. [PubMed: 301613]
4. Jorgensen WL, Tirado-Rives J. Monte Carlo vs molecular dynamics for conformational sampling. *Journal of Physical Chemistry* 1996;100:14508–14513.
5. Baker NA. Improving implicit solvent simulations: A Poisson-centric view. *Current Opinion in Structural Biology* 2005;15:137–143. [PubMed: 15837170]
6. Chen J, Brooks CL Iii, Khandogin J. Recent advances in implicit solvent-based methods for biomacromolecular simulations. *Current Opinion in Structural Biology* 2008;18:140–148. [PubMed: 18304802]
7. Onufriev, A. Chapter 7: Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. In: Wheeler, RA.; Spellmeyer, DC., editors. *Annual Reports in Computational Chemistry*. Vol. 4. Elsevier; Amsterdam: 2008. p. 125-137.
8. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*. 2nd Edition. Academic Press; San Diego, London: 2002.
9. Binder, K. *Monte Carlo Simulation in Statistical Physics*. 3rd Edition. Springer-Verlag; Berlin, Heidelberg, New York: 1997.
10. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids*. Oxford University Press; Oxford, New York: 1989.

11. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 1953;21:1087–1092.
12. Borovinskiy AL, Grosberg AY. Design of toy proteins capable of rearranging conformations in a mechanical fashion. *Journal of Chemical Physics* 2003;118:5201–5212.
13. Mazonko, GA. *Equilibrium Statistical Mechanics*. Wiley-Interscience; New York: 2000.
14. Jorgensen WL, Tirado-Rives J. Molecular modeling of organic and biomacromolecular systems using BOSS and MCPRO. *Journal of Computational Chemistry* 2005;26:1689–1700. [PubMed: 16200637]
15. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 2008;4:435–447.
16. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 2005;26:1781–1802. [PubMed: 16222654]
17. Ponder, JW. TINKER - Software Tools for Molecular Design. 4.2 ed.2004. <http://dasher.wustl.edu/tinker>
18. Case, DA.; Darden, TA.; T.E. Cheatham, I.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Crowley, M.; Walker, RC.; Zhang, W.; Merz, KM.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Kollman, PA. AMBER 10. University of California; San Francisco: 2008.
19. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 1983;4:187–217.
20. Hu J, Ma A, Dinner AR. Monte carlo simulations of biomacromolecules: The MC module in CHARMM. *Journal of Computational Chemistry* 2006;27:203–216. [PubMed: 16323162]
21. Martin, MG. MCCS Towhee. 2008. <http://towhee.sourceforge.net/>
22. Irbäck A, Mohanty S. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *Journal of Computational Chemistry* 2006;27:1548–1555. [PubMed: 16847934]
23. Kutteh, R.; Straatsma, TP. Molecular Dynamics with General Holonomic Constraints and Application to Internal Coordinate Constraints. In: Lipkowitz, LD.; Boyd, DB., editors. *Reviews in Computational Chemistry*. Vol. 12. Wiley; New York: 1998. p. 75-136.
24. Patriciu A, Chirikjian GS, Pappu RV. Analysis of the conformational dependence of mass-metric tensor determinants in serial polymers with constraints. *Journal of Chemical Physics* 2004;121:12708–12720. [PubMed: 15606297]
25. Fixman M. Classical Statistical Mechanics of Constraints: A Theorem and Application to Polymers. *Proceedings of the National Academy of Sciences of the United States of America* 1974;71:3050–3053. [PubMed: 16578718]
26. Ulmschneider JP, Ulmschneider MB, Di Nola A. Monte carlo vs molecular dynamics for all-atom polypeptide folding simulations. *Journal of Physical Chemistry B* 2006;110:16733–16742.
27. Cahill S, Cahill M, Cahill K. On the kinematics of protein folding. *Journal of Computational Chemistry* 2003;24:1364–1370. [PubMed: 12827678]
28. Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 1981;14:325–332.
29. Fitzgerald JE, Jha AK, Sosnick TR, Freed KF. Polypeptide motions are dominated by peptide group oscillations resulting from dihedral angle correlations between nearest neighbors. *Biochemistry* 2007;46:669–682. [PubMed: 17223689]
30. Vitalis A, Pappu RV. ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry* 2009;30:673–699. [PubMed: 18506808]
31. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B* 2001;105:6474–6487.
32. Skeel RD, Izaguirre JA. An impulse integrator for Langevin dynamics. *Molecular Physics* 2002;100:3885–3891.



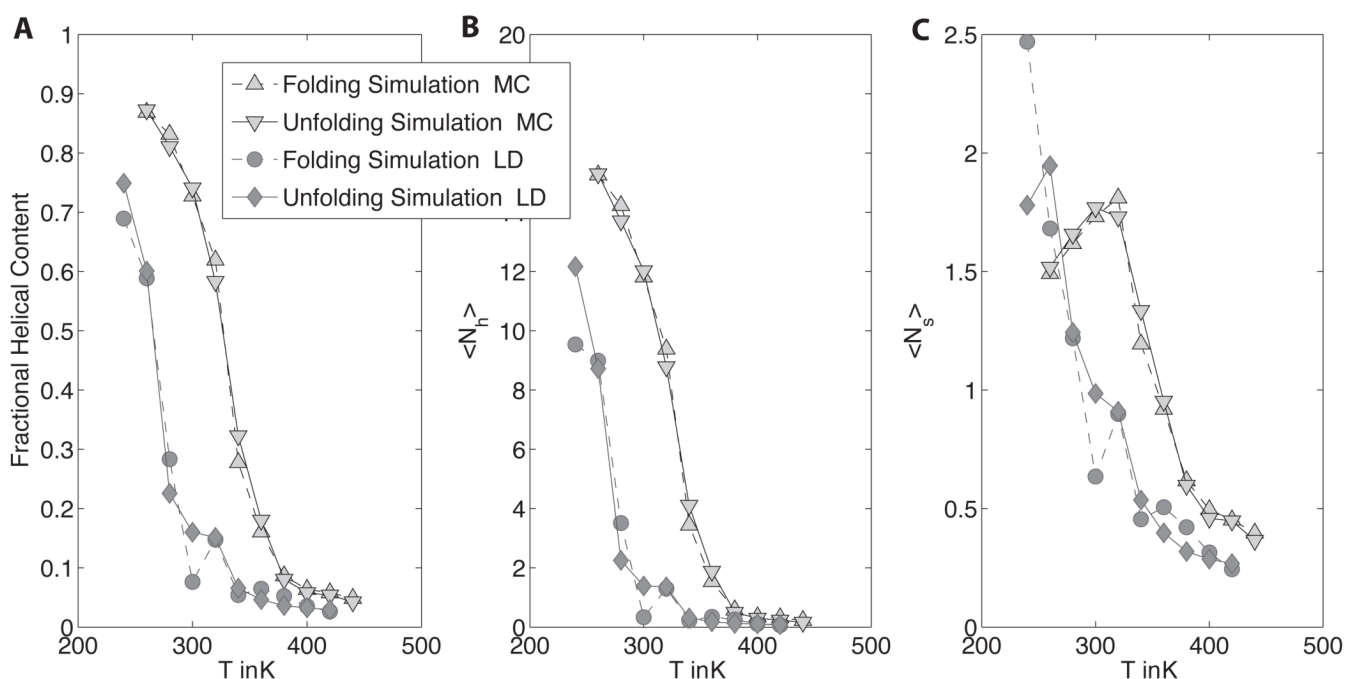
33. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *Journal of Computational Chemistry* 2003;24:1999–2012. [PubMed: 14531054]
34. Foloppe N, MacKerell AD Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *Journal of Computational Chemistry* 2000;21:86–104.
35. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kucera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE Iii, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kucera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* 1998;102:3586–3616.
36. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomacromolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* 2004;25:1656–1676. [PubMed: 15264259]
37. Darden T, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 1993;98:10089–10092.
38. Mallik B, Masunov A, Lazaridis T. Distance and exposure dependent effective dielectric function. *Journal of Computational Chemistry* 2002;23:1090–1099. [PubMed: 12116395]
39. Michel J, Taylor RD, Essex JW. Efficient generalized born models for Monte Carlo simulations. *Journal of Chemical Theory and Computation* 2006;2:732–739.
40. Nilmeier J, Jacobson M. Multiscale Monte Carlo sampling of protein sidechains: Application to binding pocket flexibility. *Journal of Chemical Theory and Computation* 2008;4:835–846. [PubMed: 19119325]
41. Gelb LD. Monte Carlo simulations using sampling from an approximate potential. *Journal of Chemical Physics* 2003;118:7747–7750.
42. Hetényi B, Bernacki K, Berne BJ. Multiple “time step” Monte Carlo. *Journal of Chemical Physics* 2002;117:8203–8207.
43. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 1995;103:8577–8593.
44. Bernacki K, Hetényi B, Berne BJ. Multiple “time step” Monte Carlo simulations: Application to charged systems with Ewald summation. *Journal of Chemical Physics* 2004;121:44–50. [PubMed: 15260521]
45. Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* 1921;369:253–287.
46. Brush SG, Sahlin HL, Teller E. Monte-Carlo Study of a One-Component Plasma. I. *J. Chem. Phys* 1966;45:2102–2118.
47. Hansen JP. Statistical mechanics of dense ionized matter. I. Equilibrium properties of the classical one-component plasma. *Physical Review A* 1973;8:3096–3109.
48. Adams DJ, Dubey GS. Taming the Ewald sum in the computer simulation of charged systems. *Journal of Computational Physics* 1987;72:156–176.
49. Toukmaji AY, Board JA Jr. Ewald summation techniques in perspective: A survey. *Computer Physics Communications* 1996;95:73–92.
50. Tironi IG, Sperb R, Smith PE, Van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics* 1995;102:5451–5459.
51. Hockney, RW.; Eastwood, JW. *Computer Simulations Using Particles*. McGraw-Hill; New York: 1981.
52. Verlet L. Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* 1967;159:98–103.
53. Murthy VL, Srinivasan R, Draper DE, Rose GD. A complete conformational map for RNA. *Journal of Molecular Biology* 1999;291:313–327. [PubMed: 10438623]
54. Lal M. ‘Monte Carlo’ computer simulation of chain molecules. I. *Molecular Physics* 1969;17:57–64.

55. Dunbrack RL Jr. Rotamer libraries in the 21st century. *Current Opinion in Structural Biology* 2002;12:431–440. [PubMed: 12163064]
56. Ho BK, Coutsias EA, Seok C, Dill KA. The flexibility in the proline ring couples to the protein backbone. *Protein Science* 2005;14:1011–1018. [PubMed: 15772308]
57. Vitagliano L, Berisio R, Mastrangelo A, Mazzarella L, Zagari A. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Science* 2001;10:2627–2632. [PubMed: 11714932]
58. Saenger, W. *Principles of Nucleic Acid Structure*. 1st Edition. Springer; New York, Berlin, Heidelberg: 1984.
59. Sklenar H, Wüstner D, Rohs R. Using internal and collective variables in Monte Carlo simulations of nucleic acid structures: Chain breakage/closure algorithm and associated Jacobians. *Journal of Computational Chemistry* 2006;27:309–315. [PubMed: 16355439]
60. Binder K, Baschnagel J, Paul W. Glass transition of polymer melts: Test of theoretical concepts by computer simulation. *Progress in Polymer Science (Oxford)* 2003;28:115–172.
61. Tanaka M, Grosberg AY, Tanaka T. Molecular dynamics simulations of polyampholytes. *Langmuir* 1999;15:4052–4055.
62. Takada S, Wolynes PG. Glassy dynamics of random heteropolymers. *Progress of Theoretical Physics Supplement* 1997:49–52.
63. Vitalis A, Wang X, Pappu RV. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophysical Journal* 2007;93:1923–1937. [PubMed: 17526581]
64. Frisch T, Verga A. Slow relaxation and solvent effects in the collapse of a polymer. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2002;66:041807.
65. Ziv G, Thirumalai D, Haran G. Collapse transition in proteins. *Physical Chemistry Chemical Physics* 2009;11:83–93. [PubMed: 19081910]
66. Straub JE, Thirumalai D. Exploring the energy landscape in proteins. *Proceedings of the National Academy of Sciences of the United States of America* 1993;90:809–813. [PubMed: 8430090]
67. Van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, Van Der Vegt NFA, Yu HB. Biomacromolecular modeling: Goals, problems, perspectives. *Angewandte Chemie - International Edition* 2006;45:4064–4092.
68. Hess B. Convergence of sampling in protein simulations. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2002;65:031910.
69. Frickenhaus S, Kannan S, Zacharias M. Efficient evaluation of sampling quality of molecular dynamics simulations by clustering of dihedral torsion angles and sammon mapping. *Journal of Computational Chemistry* 2009;30:479–492. [PubMed: 18680215]
70. Son WJ, Jang S, Shin S. A simple method of estimating sampling consistency based on free energy map distance. *Journal of Molecular Graphics and Modelling* 2008;27:321–325. [PubMed: 18619879]
71. Monticelli L, Sorin EJ, Tieleman DP, Pande VS, Colombo G. Molecular simulation of multistate peptide dynamics: A comparison between microsecond timescale sampling and multiple shorter trajectories. *Journal of Computational Chemistry* 2008;29:1740–1752. [PubMed: 18307167]
72. Karney CFF. Quaternions in molecular modeling. *Journal of Molecular Graphics and Modelling* 2007;25:595–604. [PubMed: 16777449]
73. Franklin J, Doniach S. Adaptive time stepping in biomacromolecular dynamics. *The Journal of chemical physics* 2005;123:124909. [PubMed: 16392529]
74. Barth E, Leimkuhler B, Reich S. Time-reversible variable-stepsize integrator for constrained dynamics. *SIAM Journal of Scientific Computing* 1999;21:1027–1044.
75. Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD. Langevin stabilization of molecular dynamics. *Journal of Chemical Physics* 2001;114:2090–2098.
76. Vitalis A, Wang X, Pappu RV. Atomistic Simulations of the Effects of Polyglutamine Chain Length and Solvent Quality on Conformational Equilibria and Spontaneous Homodimerization. *Journal of Molecular Biology* 2008;384:279–297. [PubMed: 18824003]

77. Sokal, AD. Monte Carlo Methods for the Self-Avoiding Walk. In: Binder, K., editor. Monte Carlo and Molecular Dynamics Simulations in Polymer Science. Oxford University Press; New York, Oxford: 1995. p. 47-124.
78. Podtelezchnikov AA, Wild DL. Exhaustive Metropolis Monte Carlo sampling and analysis of polyalanine conformations adopted under the influence of hydrogen bonds. *Proteins: Structure, Function and Genetics* 2005;61:94-104.
79. Go N, Scheraga HA. Ring Closure and Local Conformational Deformations of Chain Molecules. *Macromolecules* 1970;3:178-187.
80. Brucoleri RE, Karplus M. Chain closure with bond angle variations. *Macromolecules* 1985;18:2767-2773.
81. Ulmschneider JP, Jorgensen WL. Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *Journal of Chemical Physics* 2003;118:4261-4271.
82. Dodd LR, Boone TD, Theodorou DN. A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Molecular Physics* 1993;78:961-996.
83. Hoffmann D, Knapp EW. Polypeptide folding with off-lattice Monte Carlo dynamics: The method. *European Biophysics Journal* 1996;24:387-403.
84. Dinner AR. Local Deformations of Polymers with Nonplanar Rigid Main-Chain Internal Coordinates. *Journal of Computational Chemistry* 2000;21:1132-1144.
85. Mezei M. Efficient Monte Carlo sampling for long molecular chains using local moves, tested on a solvated lipid bilayer. *Journal of Chemical Physics* 2003;118:3874-3879.
86. Favrin G, Irbäck A, Sjunnesson F. Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space. *Journal of Chemical Physics* 2001;114:8154-8158.
87. Coutsias EA, Seok C, Jacobson MP, Dill KA. A Kinematic View of Loop Closure. *Journal of Computational Chemistry* 2004;25:510-528. [PubMed: 14735570]
88. Kolodny R, Guibas L, Levitt M, Koehl P. Inverse kinematics in biology: The protein loop closure problem. *International Journal of Robotics Research* 2005;24:151-163.
89. Swendsen RH, Wang JS. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* 1987;58:86-88. [PubMed: 10034599]
90. Wolff U. Collective Monte Carlo Updating for Spin Systems. *Physical Review Letters* 1989;62:361-364. [PubMed: 10040213]
91. Luijten, E. Introduction to cluster Monte Carlo algorithms. In: Ferrario, M.; Ciccotti, G.; Binder, K., editors. *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1, Volume 703*. Springer; Berlin, Heidelberg, New York: 2006. p. 13-38.
92. Liu J, Luijten E. Rejection-Free Geometric Cluster Algorithm for Complex Fluids. *Physical Review Letters* 2004;92:035504. [PubMed: 14753886]
93. Dress C, Krauth W. Cluster algorithm for hard spheres and related systems. *Journal of Physics A: General Physics* 1995;28:L597-L601.
94. Whitelam S, Geissler PL. Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *Journal of Chemical Physics* 2007;127:154101. [PubMed: 17949126]
95. Bhattacharyay A, Troisi A. Self-assembly of sparsely distributed molecules: An efficient cluster algorithm. *Chemical Physics Letters* 2008;458:210-213.
96. Liu J, Luijten E. Generalized geometric cluster algorithm for fluid simulation. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2005;71:066701.
97. Maggs AC. Multiscale Monte Carlo algorithm for simple fluids. *Physical Review Letters* 2006;97:197802. [PubMed: 17155659]
98. Pangali C, Rao M, Berne BJ. On a novel Monte Carlo scheme for simulating water and aqueous solutions. *Chemical Physics Letters* 1978;55:413-417.
99. Rossky PJ, Doll JD, Friedman HL. Brownian dynamics as smart Monte Carlo simulation. *The Journal of Chemical Physics* 1978;69:4628-4633.
100. Heyes DM, Branka AC. Monte Carlo as Brownian dynamics. *Molecular Physics* 1998;94:447-454.
101. Kotelyanskii MJ, Suter UW. A dynamic Monte Carlo method suitable for molecular simulations. *The Journal of Chemical Physics* 1992;96:5383-5388.

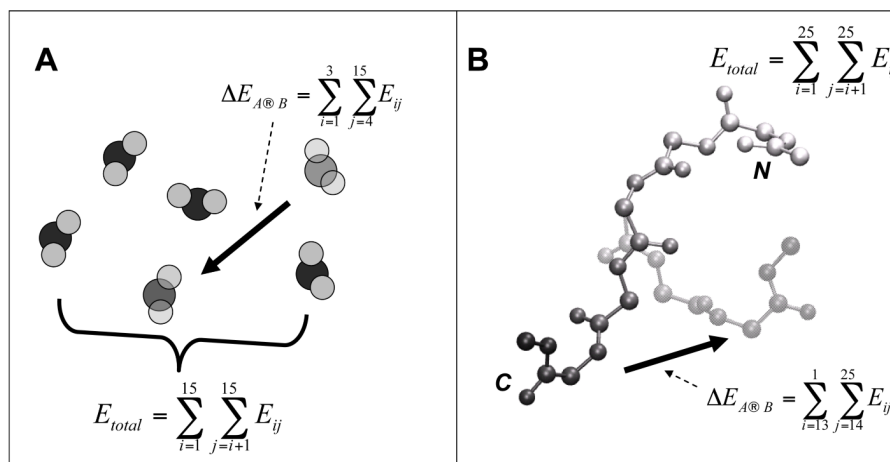
102. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Physics Letters B* 1987;195:216–222.
103. Kennedy AD, Pendleton B. Acceptances and autocorrelations in hybrid Monte Carlo. *Nuclear Physics B (Proceedings Supplements)* 1991;20:118–121.
104. Mackenzie PB. An improved hybrid Monte Carlo method. *Physics Letters B* 1989;226:369–371.
105. Faller R, De Pablo JJ. Constant pressure hybrid molecular dynamics–Monte Carlo simulations. *Journal of Chemical Physics* 2002;116:55–59.
106. Izaguirre JA, Hampton SS. Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules. *Journal of Computational Physics* 2004;200:581–604.
107. Akhmatskaya E, Bou-Rabee N, Reich S. A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *Journal of Computational Physics* 2009;228:2256–2265.
108. Kennedy AD. Higher order hybrid Monte Carlo. *Nuclear Physics B (Proceedings Supplements)* 1989;9:457–462.
109. Berendsen HJC, Postma JPM, Van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 1984;81:3684–3690.
110. Kremer K, Binder K. Monte Carlo simulation of lattice models for macromolecules. *Computer Physics Reports* 1988;7:259–310.
111. Nidras PP, Brak R. New Monte Carlo algorithms for interacting self-avoiding walks. *Journal of Physics A: Mathematical and General* 1997;30:1457–1469.
112. Siepmann JI, Frenkel D. Configurational bias Monte Carlo: A new sampling scheme for flexible chains. *Molecular Physics* 1992;75:59–70.
113. O'Toole EM, Panagiotopoulos AZ. Monte Carlo simulation of folding transitions of simple model proteins using a chain growth algorithm. *The Journal of Chemical Physics* 1992;97:8644–8652.
114. Frenkel D, Mooij GCAM, Smit B. Novel scheme to study structural and thermal properties of continuously deformable molecules. *Journal of Physics: Condensed Matter* 1992;4:3053–3076.
115. Vendruscolo M. Modified configurational bias Monte Carlo method for simulation of polymer systems. *Journal of Chemical Physics* 1997;106:2970–2976.
116. Vlught TJH, Martin MG, Smit B, Siepmann JI, Krishna R. Improving the efficiency of the configurational-bias Monte Carlo algorithm. *Molecular Physics* 1998;94:727–733.
117. Zhang J, Kou SC, Liu JS. Biopolymer structure simulation and optimization via fragment regrowth Monte Carlo. *Journal of Chemical Physics* 2007;126:225101. [PubMed: 17581081]
118. Escobedo FA, De Pablo JJ. Extended continuum configurational bias Monte Carlo methods for simulation of flexible molecules. *The Journal of Chemical Physics* 1995;102:2636–2652.
119. Houdayer J. The wormhole move: A new algorithm for polymer simulations. *Journal of Chemical Physics* 2002;116:1783–1787.
120. Laso M, Karayiannis NC, Müller M. Min-map bias Monte Carlo for chain molecules: Biased Monte Carlo sampling based on bijective minimum-to-minimum mapping. *Journal of Chemical Physics* 2006;125:164108. [PubMed: 17092064]
121. Grassberger P. Pruned-enriched Rosenbluth method: Simulations of  $\theta$  polymers of chain length up to 1 000 000. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 1997;56:3682–3693.
122. Consta S, Vlught TJH, Hoeth JW, Smit B, Frenkel D. Recoil growth algorithm for chain molecules with continuous interactions. *Molecular Physics* 1999;97:1243–1254.
123. Prentiss MC, Wales DJ, Wolynes PG. Protein structure prediction using basin-hopping. *Journal of Chemical Physics* 2008;128:225106. [PubMed: 18554063]
124. Sugita, Y.; Mitsutake, A.; Okamoto, Y. Generalized-ensemble algorithms for protein folding simulations. In: Janke, W., editor. *Rugged Free Energy Landscapes*. Vol. 736. Springer; Berlin, Heidelberg, New York: 2008. p. 369-407.
125. Frantz DD, Freeman DL, Doll JD. Reducing quasi-ergodic behavior in Monte Carlo simulations by J-walking: Applications to atomic clusters. *The Journal of Chemical Physics* 1990;93:2769–2784.
126. Zhou R, Berne BJ. Smart walking: A new method for Boltzmann sampling of protein conformations. *Journal of Chemical Physics* 1997;107:9185–9196.

127. Xu H, Berne BJ. Multicanonical jump walking: A method for efficiently sampling rough energy landscapes. *Journal of Chemical Physics* 1999;110:10299–10306.
128. Brown S, Head-Gordon T. Cool walking: A new Markov chain Monte Carlo sampling method. *Journal of Computational Chemistry* 2003;24:68–76. [PubMed: 12483676]
129. Andricioaei I, Straub JE, Voter AF. Smart darting Monte Carlo. *Journal of Chemical Physics* 2001;114:6994–7000.
130. Nigra P, Freeman DL, Doll JD. Combining smart darting with parallel tempering using Eckart space: Application to Lennard-Jones clusters. *Journal of Chemical Physics* 2005;122:1.
131. Wang F, Landau DP. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters* 2001;86:2050. [PubMed: 11289852]
132. Shell MS, Debenedetti PG, Panagiotopoulos AZ. Generalization of the Wang-Landau method for off-lattice simulations. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2002;66:056703.
133. Wüst T, Landau DP. The HP model of protein folding: A challenging testing ground for Wang-Landau sampling. *Computer Physics Communications* 2008;179:124–127.
134. Thomas GL, Sessions RB, Parker MJ. Density guided importance sampling: Application to a reduced model of protein folding. *Bioinformatics* 2005;21:2839–2843. [PubMed: 15802285]
135. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of Molecular Biology* 2001;308:79–95. [PubMed: 11302709]
136. Shimada J, Shakhnovich EI. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:11175–11180. [PubMed: 12165568]
137. Nivón LG, Shakhnovich EI. All-atom Monte Carlo simulation of GCAA RNA folding. *Journal of Molecular Biology* 2004;344:29–45. [PubMed: 15504400]
138. Shental-Bechor D, Kirca S, Ben-Tal N, Haliloglu T. Monte Carlo studies of folding, dynamics, and stability in  $\alpha$ -helices. *Biophysical Journal* 2005;88:2391–2402. [PubMed: 15653741]
139. Irbäck, A. Protein folding, unfolding and aggregation studied using an all-atom model with a simplified interaction potential. In: Janke, W., editor. *Rugged Free Energy Landscapes*. Vol. 736. Springer; Berlin, Heidelberg: 2008. p. 269-291.
140. Li DW, Mohanty S, Irbäck A, Huo S. Formation and growth of oligomers: A monte carlo study of an amyloid tau fragment. *PLoS Computational Biology* 2008;4:e1000238. [PubMed: 19057640]
141. Mitternacht S, Luccioli S, Torcini A, Imparato A, Irbäck A. Changing the mechanical unfolding pathway of FnIII10 by tuning the pulling strength. *Biophysical Journal* 2009;96:429–441. [PubMed: 19167294]
142. Ulmschneider JP, Ulmschneider MB, Di Nola A. Monte Carlo folding of trans-membrane helical peptides in an implicit generalized Born membrane. *Proteins: Structure, Function and Genetics* 2007;69:297–308.
143. Ulmschneider JP, Ulmschneider MB. Folding simulations of the transmembrane Helix of virus protein U in an implicit membrane model. *Journal of Chemical Theory and Computation* 2007;3:2335–2346.
144. De Mori GMS, Colombo G, Micheletti C. Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins: Structure, Function and Genetics* 2005;58:459–471.
145. Meli M, Morra G, Colombo G. Investigating the mechanism of peptide aggregation: Insights from mixed monte carlo-molecular dynamics simulations. *Biophysical Journal* 2008;94:4414–4426. [PubMed: 18263661]

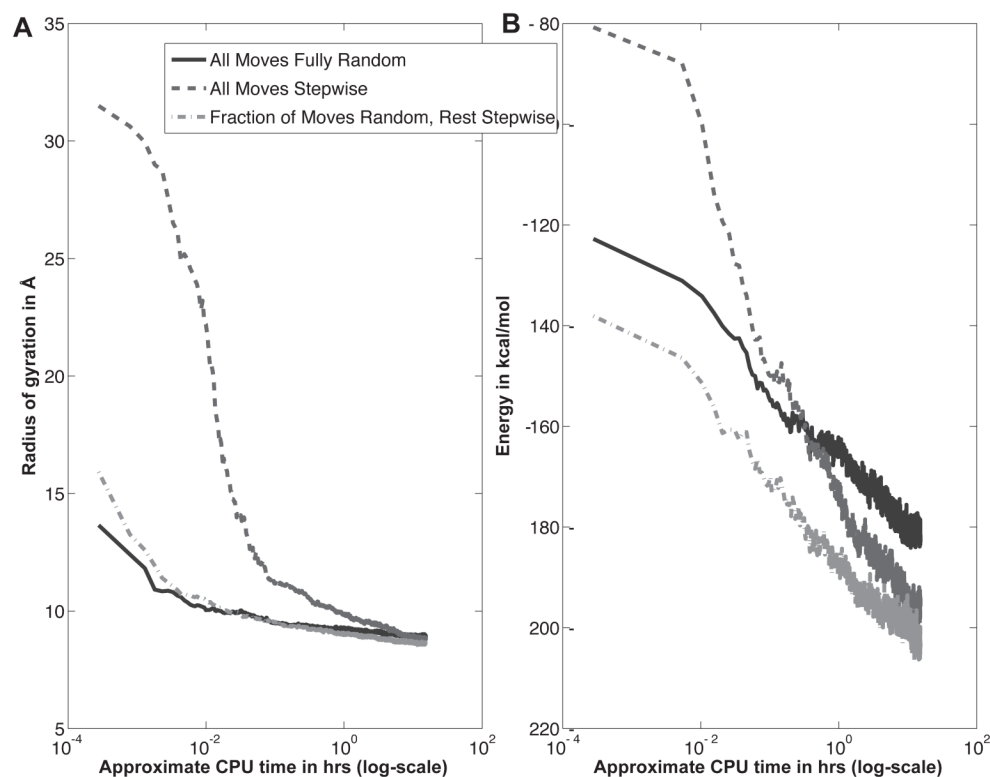


**Figure 1.**

Melting of the FS-peptide as a function of simulation temperature. This Figure is analogous to Figure 8 in [30] and the reader is referred there for details of the computation of helix parameters and the description of the MC data. Panel A shows net  $\alpha$ -helical content, Panel B the mean number of  $\alpha$ -helical hydrogen bonds, and Panel C the mean number of  $\alpha$ -helical segments. LD data were obtained using the impulse integrator of Skeel and Izaguirre [32] with uniform friction of  $5.0\text{ps}^{-1}$  and an integration timestep of 1fs. The total simulation time at each temperature was 25ns, the first 10ns of which were discarded for equilibration purposes. No constraints were used, and all bond angle and bond length parameters were adopted directly from OPLS-AA/L [31]. No torsions were employed except those keeping quasi-rigid units planar. Those included the peptide bonds and the guanidinium groups in the arginine sidechains.

**Figure 2.**

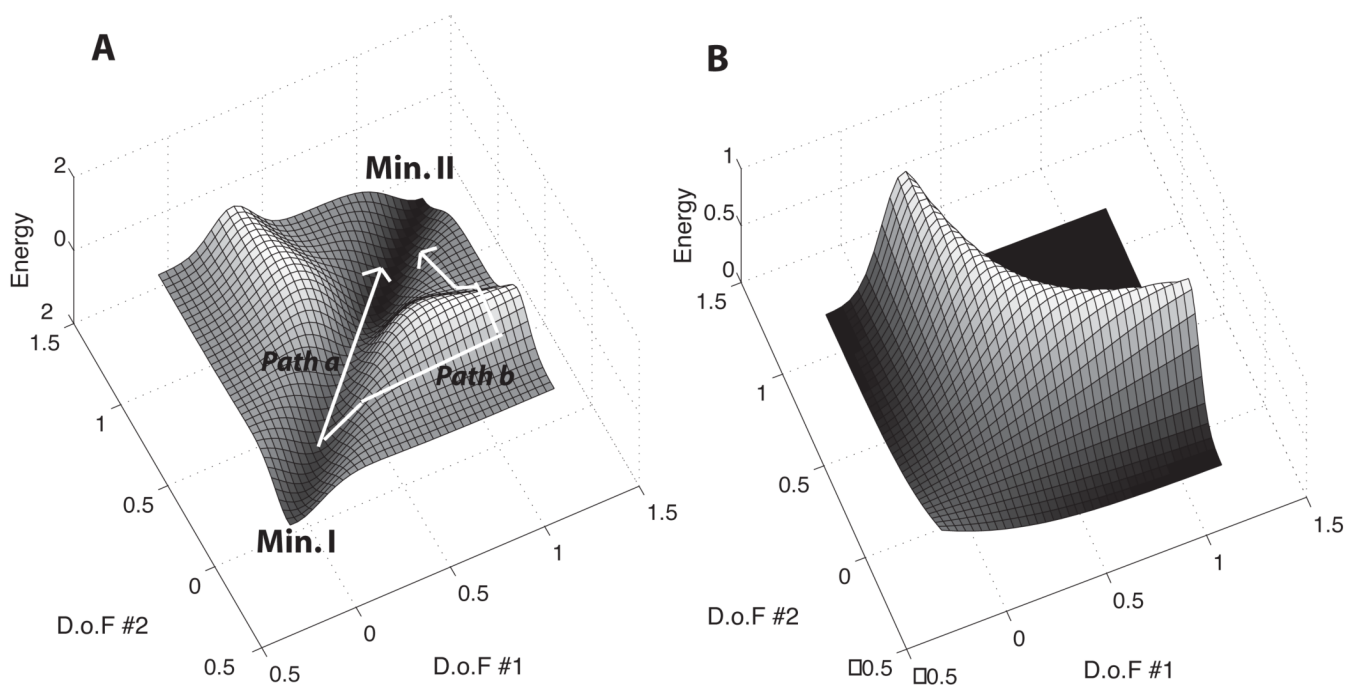
Two examples illustrating complexity of incremental energy updates in MC calculations. Panel A shows the displacement of a single three-site molecule (lighter color) in the presence of four other such molecules. For a strictly pairwise additive Hamiltonian, the only changing terms are the ones involving the moving molecule (atoms 1-3). Therefore, the resultant energy calculation only needs to encompass 36 of the 105 total pairwise site-site energies. Panel B shows the rotation (pivot) around a single dihedral angle in a molecule composed of 25 sites (assuming systematic numbering from N- to C-terminus). Upon pivoting, the two arms move relative to one another, but remain rigid internally. 156 out of the total of 300 site-site interactions need to be computed, and the efficiency scales poorly with molecule size (see text).



**Figure 3.**

The collapse of the peptide Ace-Nle<sub>30</sub>-Nme under deeply quenched poor solvent conditions monitored by both radius of gyration (Panel A) and energy relaxation (Panel B). MC simulations were performed in dihedral space. 81% of moves attempted to change  $\phi/\psi$ -angles, 9% sampled the  $\omega$ -angles, and 10% the sidechains. For the randomized case (solid line), all angles were uniformly sampled from the interval  $-180^\circ$  to  $180^\circ$  each time. For the stepwise case (dashed line), dihedral angles were perturbed uniformly by a maximum of  $10^\circ$  for  $\phi/\psi$ -moves,  $2^\circ$  for  $\omega$ -moves, and  $30^\circ$  for sidechain moves. In the mixed case (dash-dotted line), the stepwise protocol was modified to include non-local moves with fractions of 20% for  $\phi/\psi$ -moves, 10% for  $\omega$ -moves, and 30% for sidechain moves. For each of the three cases, data from 20 independent runs were combined to yields the traces shown. CPU times are approximate since stochastic variations in runtime were observed for the independent runs. Each run was comprised of  $3 \times 10^7$  steps. Error estimates are not shown in the interest of clarity, but indicated the results to be robust.





**Figure 4.** Two arbitrary potential energy surfaces in a two-dimensional coordinate space. All units are arbitrary. Panel A shows two minima connected by a path in phase space requiring correlated change in both degrees of freedom (labeled Path a). As is indicated, paths involving sequential change of the degrees of freedom encounter a large enthalpic barrier (labeled Path b). Panel B shows two minima separated by a barrier. No path with a small enthalpic barrier is available and correlated, stepwise evolution of the system is not sufficient for barrier-crossing.