Equilibrium Sampling in Biomolecular Simulations

Daniel M. Zuckerman

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213; email: ddmmzz@pitt.edu

Keywords

Abstract

Equilibrium sampling of biomolecules remains an unmet challenge after more than 30 years of atomistic simulation. Efforts to enhance sampling capability, which are reviewed here, range from the development of new algorithms to parallelization to novel uses of hardware. Special focus is placed on classifying algorithms—most of which are underpinned by a few key ideas—in order to understand their fundamental strengths and limitations. Although algorithms have proliferated, progress resulting from novel hardware use appears to be more clear-cut than from algorithms alone, due partly to the lack of widely used sampling measures.

Contents

INTRODUCTION

Why Sample?

Biomolecular behavior can be substantially characterized by the states of the system of interest—that is, by the configurational energy basins reflecting the coordinates of all constituent molecules in a system. These states might be largely internal to a single macromolecule, such as a protein, or more generally involve binding partners and their relative coordinates. But regardless of the complexity of a system, the states represent key functional configurations along with potential intermediates for transitioning among the major states. The primary goal of equilibrium computer simulation is to specify configurational states and their populations. A more complete, mechanistic description would also include kinetic properties and dynamical pathways (147). Nevertheless, even the equilibrium description provides a basic view of the range of structural motions of biomolecules and can serve more immediately practical purposes such as for ensemble docking (56, 65).

The basic algorithm of biomolecular simulation, molecular dynamics (MD) simulation, has not changed substantially since the first MD study of a protein more than 30 years ago (82). Although routine explicit-solvent MD simulations are now four or five orders of magnitude longer (i.e., $100–10^3$ ns currently), modern MD studies still appear to fall significantly short of what is needed for statistically valid equilibrium simulation (36, 38). Roughly speaking, one would like to run a simulation at least 10 times longer than the slowest important timescale in a system. Unfortunately, many biomolecular timescales exceed 1 ms, and in some cases by orders of magnitude (44).

Despite the outpouring of algorithmic ideas over the past decades, MD largely remains the tool of choice for biomolecular simulations. While this staying power partly reflects the ready availability of software packages, and perhaps some psychological inertia, it also is indicative of a simple fact: No other method can

routinely and reliably outperform MD by a significant amount.

This review employs several points of view in considering efforts to improve sampling. First, it attempts to define the equilibrium sampling problem(s) as precisely as possible, which necessarily includes discussing the quantification of sampling. Significant and substantiated progress is not possible without clear yardsticks. Second, a brief discussion of how sampling happens—i.e., how simulations generate ensembles and averages—provides a basis for understanding numerous methods. Third, the review attempts to examine a fairly wide array of modern algorithms—most of which are underpinned by a surprisingly small number of key ideas. The algorithms and studies are too numerous to be reviewed on a case-by-case basis, but a bird's-eye view is informative; more focused attention is paid to apparently conflicting statements about the replica exchange approach, however. Fourth, special emphasis is be placed on novel uses of hardware—graphics processing units (GPUs), random-access memory (RAM), and special central processing units (CPUs); this new front in sampling efforts has yielded rather clean results in some cases. Using hardware in new ways necessarily involves substantial algorithmic efforts.

This review is limited. At least a full volume would be required to comprehensively describe sampling methods and results for biomolecular systems. This review therefore aims to primarily catalog key ideas and principles of sampling, along with enough references for the reader to delve more deeply into areas of interest. The author apologizes for the numerous studies that have not been mentioned because of space limitations or because he was not aware of them. Other review articles (e.g., 4, 17) and books on simulation methods (2, 32, 62, 105) are available.

WHAT IS THE SAMPLING PROBLEM?

Although scientists actively working in the field of biomolecular simulation may assume the essential meaning of the sampling problem is universally accepted, a survey of the literature indicates that several somewhat different interpretations are implicitly assumed. An individual's understanding of the sampling problem and, accordingly, success or failure in addressing the problem surely will dictate the choice of methodology.

A Simple View of the Equilibrium Sampling Problem

Equilibrium sampling at constant temperature and fixed volume can be concisely identified with the generation of full-system configurations, $\mathbf{x}$, distributed according to the Boltzmann-factor distribution:

$$\rho(\mathbf{x}) \propto \exp[-U(\mathbf{x})/k_B T], \qquad 1.$$

where ρ is the probability density function, U is the potential energy function, k_B is Boltzmann's constant, and T is the absolute temperature in Kelvin units. The coordinates $\mathbf{x}$ refer (in the classical picture assumed here) to the coordinates of every atom in the system, including solute and solvent. Note that other thermodynamic conditions, such as constant pressure or constant chemical potential, require additional energy-like terms in the Boltzmann factor (147), but here we shall consider only the distribution of Equation 1 for clarity and simplicity.

Performing equilibrium sampling requires access to all regions of configuration space, or at least to those regions with significant populations, and that configurations have the correct relative probabilities. This point is further discussed below.

Ideal Sampling as a Reference Point

In practice, it is nearly impossible to generate completely independent and identically distributed (i.i.d.) configurations—obeying Equation 1—for biomolecules. Both dynamical and nondynamical methods tend to produce correlated samples (38, 41, 73, 86, 141). Nevertheless, such ideal sampling is a highly useful

MD: molecular dynamics

GPU: graphics processing unit
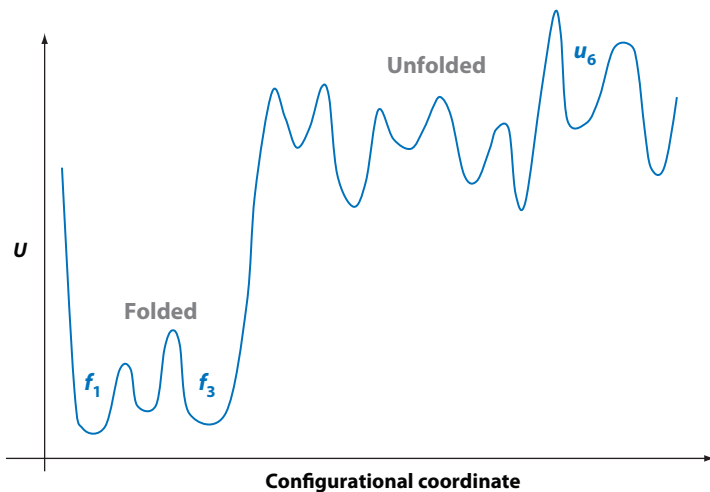
CPU: central processing unit

Figure 1

A schematic energy landscape of a protein. Both the folded and unfolded states generally can be expected to consist of multiple substates, f_i and u_i, respectively. Transitions among substates themselves could be slow, compared with simulation timescales. The sampling problem is sometimes construed to involve generating configurations from both folded and unfolded states; the text discusses when this view is justified.

reference point for disentangling complications arising in practical simulation methods that generate correlated configurations.

Consider a hypothetical example in which an equilibrium ensemble of N i.i.d. configurations has been generated according to Equation 1 for a biomolecule with a complex landscape. Regions of configuration space with probability $p > 1/N$ are likely to be represented in the ensemble, and this is true regardless of kinetic barriers among states. Conversely, regions with $p < 1/N$ are not likely to be represented, whether or not the region is interesting. For instance, the unfolded state might be an appropriate part of a size N ensemble, but this will depend on the system (i.e., on the free energy difference between folded and unfolded states) and on N (see **Figure 1** and the discussion below).

Goals and Ambiguities in Equilibrium Sampling

The question, "What should be achieved in equilibrium sampling?" can be divided into at least three subquestions. First, for a given problem (defined by the specific system and a specified initial condition), there is the question of defining success—When is sampling sufficient? Usually, the goal of equilibrium simulation is to calculate observables of interest to a level of precision sufficient to draw physical conclusions. Yet the variables of interest vary from study to study. It therefore seems more reasonable to set a goal of generating an ensemble of configurations from which arbitrary observables can be calculated. In rough terms, let us therefore assume that the goal is to calculate N^{eff} effectively independent configurations, where presumably it is desirable to achieve $N^{eff} > 10$. Larger values may often be necessary, however, because fractional uncertainty will vary as $1/\sqrt{N^{eff}}$ for slower-to-converge properties such as state populations. This issue is addressed further below.

The other questions are probably the most pressing, starting with, "What type of system is to be sampled?" A basic distinction that seems to appear implicitly in the literature is whether the unfolded state is important in the system. In this context, importance would usually be defined by whether a system exhibits a significant equilibrium unfolded fraction—say, roughly equal probabilities p(folded) $\sim$ p(unfolded). Indeed, some evaluations of sampling methods have specifically targeted such a balance (48, 102), e.g., by adjusting the temperature. A third question concerns initial conditions, which sometimes correspond to the unfolded state. When will the initial conditions be important? This issue is addressed separately below.

These questions arise in the first place because all available practical sampling methods for biomolecules are imperfect in the sense of producing correlated configurations. In ideal equilibrium sampling, by contrast, the sufficiency of sampling is fully quantified in terms of the number N of independent configurations and can be assessed objectively in terms of observables of interest. The notion of an initial condition is irrelevant in ideal sampling because there are no correlations. Also, the values of p(folded) and p(unfolded) have no effect on sampling quality, which is fully embodied in N;

instead, it is N that determines which states are represented in the ensemble and the appropriate frequencies of occurrence.

Issues surrounding initial conditions and the unfolded state. A typical simulation of a protein can be described as one intended to sample the folded (presumably functional) state, which may consist of a number of substates as in **Figure 1**. One or more of the folded substates may be partially disordered. So long as the simulation is initiated in or near one of the nominally folded substates, the resources required to sample it will not depend significantly on the initial conditions. Rather, timescales for transitioning among all folded substates will dominate the computing cost. In a dynamical simulation, after all, it is the transitions among substates that yield estimates for substate populations and hence the equilibrium ensemble.

Sensitivity to initial conditions can become important when the unfolded state is involved. Even when the goal is to sample the equilibrium ensemble of an overwhelmingly folded protein, the initial condition can come into play if it introduces an anomalously long timescale. Consider beginning an equilibrium simulation of the landscape of **Figure 1** in the unfolded substate labeled u_6. Depending on details of the system, the time required to find the folded state could easily exceed the time for sampling folded substates. Thus, such a simulation must first solve a search problem before sampling can begin.

Aside from the issue of initial conditions, the relative importance of the unfolded state can be considered in the context of the folding free energy, which is defined on the basis of the ratio of probabilities for unfolded and folded states:

$$p(\text{folded})/p(\text{unfolded}) = \exp(-\Delta G_{\text{fold}}/k_B T).$$
$$2.$$

Consider the value $\Delta G_{\text{fold}} \sim -3\,\text{kcal/mol} \sim -5 k_B T$, which implies the unfolded state is occupied 1% of the time or less. In ensembles with fewer than 100 independent configurations ($N^{\text{eff}} < 100$), such an unfolded state typically should not be represented. In practical

terms, because key folded-state timescales often exceed microseconds or even milliseconds, it is unlikely that a simulation of a protein in folding conditions in the current era will contain a sufficient number of independent configurations for the unfolded state to be represented.

Note that we have assumed it is straightforward to distinguish folded and unfolded configurations. In reality, there may be a spectrum of disordered states in many systems and hence ambiguity in defining folded and unfolded states.

SAMPLING BASICS: MECHANISM, TIMESCALE, AND COST

Transitions Provide the Key Information in Sampling

The problem of equilibrium sampling can be summarized as determining the metastable states and their relative populations, but how are populations actually determined by the simulation? In almost every practical algorithm, the information comes from transitions among the states. By switching back and forth between pairs of states, with dwell times between transitions acting as proxies for rate constants k, a dynamical simulation gathers information about relative populations based on the equilibrium balance relation $p_i k_{ij} = p_j k_{ji}$ (147). With more transitions, the error in the population estimates p_i decreases (102). Without transitions, most algorithms have no information about relative populations. Consider the case of two independent simulations started from different states that exhibit no transitions: Determining the populations of the states then requires more advanced analysis (136) that often may not be practical.

Most advanced algorithms—such as the varieties of exchange simulation—also require ordinary transitions to gather population information. That is, transitions among states within individual continuous trajectories are required to obtain sampling, as has been discussed previously (38, 102, 143). Any algorithm that uses dynamical trajectories as a component can be

expected to require transitions among metastable states in those trajectories.

It must also be pointed out that transitions themselves do not necessarily equate to good sampling. An example is when a high temperature is used to aid sampling at a lower temperature. Although transitions are necessary, they are not sufficient because of the overlap issue: The sampled configurations ultimately must be important in the targeted (e.g., lower temperature) ensemble after they are reweighted (72, 109).

Timescales of Sampling

Much of the foregoing discussion can be summed up based on two key timescales. (Even if a sampling method is not fully dynamical, analogous quantities can be defined in terms of computing times.) The timescale that typically is limiting for simulations can be denoted t^*_{corr}, which denotes the longest important equilibrium correlation time. Roughly, this is the time required to explore all the important parts of configuration space once (73)—starting from any well-populated (sub)state. For sampling the landscape of **Figure 1** under folding conditions, t^*_{corr} would be the time to visit all folded substates starting from any of the folded basins. However, as also discussed above, initial conditions may play an important role, and thus we can define t_{init} as the equilibration time (or burn-in time in Monte Carlo lingo). This is not an equilibrium timescale, but the time necessary to relax from a particular nonequilibrium initial condition to equilibrium. Thus, t_{init} is specific to the initial configuration of each simulation. The landscape of **Figure 1** under folding conditions suggests $t_{\text{init}} \ll t^*_{\text{corr}}$ for a simulation started from a folded substate, but it is possible to have significant values—even $t_{\text{init}} > t^*_{\text{corr}}$—when starting from an unfolded state.

Factors Contributing to Sampling Cost

A little thought about the ingredients of sampling can help one understand successes and failures of various efforts and, more importantly, aid in planning future research. The most important ingredient to understand is single-trajectory sampling, i.e., the class of dynamical methods including MD, single-Markov-chain Monte Carlo (MC), or any other method that generates an ensemble as a sequentially correlated list of configurations. Single-trajectory methods tend to be the engine underlying traditional and advanced methods.

The sampling cost of a trajectory method can be divided into two intuitive factors:

Trajectory sampling cost

= (Cost per trajectory step)

×(Number of steps needed for sampling)

∼ (Cost of energy call)

×(Roughness of energy landscape).   3.

In words, a given algorithm requires a certain number of steps to achieve good sampling (e.g., $N^{\text{eff}} \gg 1$), and the total cost of generating such a trajectory is simply the total cost per step multiplied by the required number of steps. Although this decomposition is trivial, it is informative. For example, if one wants to use a single-trajectory algorithm (e.g., MD or similar) where a step corresponds roughly to 10^{-15} s, then 10^9–10^{12} steps are required to reach the microsecond to millisecond range. Therefore, unless the cost of a step can be reduced by several orders of magnitude, MD and similar methods likely cannot achieve sampling with typical current resources. In a modified landscape (e.g., different temperature or model), the roughness may be reduced, but again that reduction should be several orders of magnitude or else be accompanied by a compensating reduction in the cost per step.

Among noteworthy examples, increasing temperature decreases roughness but not the cost per step, whereas changing a model (i.e., modifying the potential energy function U) can affect both the roughness and the step cost. Both strategies sample a different distribution, which may be nontrivial to convert to the targeted ensemble. This type of understanding, building on Equation 3, can aid in the analysis of

Correlation time: the time that must elapse, on average, between evaluations of a specific observable for the values to be statistically independent

MC: Monte Carlo

multilevel (e.g., exchange) simulations, which require good sampling at some level (143, 148) by single continuous trajectories.

QUANTITATIVE ASSESSMENT OF SAMPLING

Although the challenge of sampling biomolecular systems has been widely recognized for years, and although new algorithms are regularly proposed, systematic quantification of sampling has often not received sufficient attention. Most importantly, there does not seem to be a widely accepted (and widely used) yardstick for quantifying the effectiveness of sampling procedures. Given the lack of a broadly accepted measure, it is tempting for individual investigators to report measures that cast the most favorable light on their results.

Two key benefits would flow from a universal measure of sampling quality. As an example, assume we are able to assign an effective sample size, N^{eff}, to any ensemble generated in a simulation; N^{eff} would characterize the number of statistically independent configurations generated. Sampling efficiency could then be quantified by the CPU time (or number of cycles) required per independent sample. The first benefit is that we would no longer be able to fool ourselves as to the efficiency of a method based on qualitative evidence or perhaps the elegance of a method. Thus, the field would be pushed harder to focus on a bottom-line measure. The second benefit is that demonstrating the importance of a new algorithm would be more straightforward. It would no longer be necessary to directly compare a new method with, perhaps, a competitor's approach requiring subtle optimization. Instead, each method could be compared with a standard (e.g., MD), removing ambiguities in the outcome. It would still be important to test a method on a number of systems (e.g., small and large, stiff and flexible, implicitly and explicitly solvated), but the results would be quantitative and readily verified by other groups. On a related note, there has been a proposal to organize a sampling contest or challenge event to allow head-to-head

comparison of different methodologies (B.R. Brooks, unpublished communication).

Numerous ideas for assessing sampling have been proposed over the years (36, 38, 41, 73, 86, 88, 114, 116, 141), but it is important to divide such proposals into absolute and relative measures. Absolute measures attempt to give a binary indication of whether convergence has been achieved, whereas relative measures estimate how much sampling has been achieved, e.g., by an effective sample size N^{eff}. The perspective of the present review is that absolute measures fail to account for the fundamental statistical picture underlying the sampling problem. To understand why, note that any measured observable has an uncertainty associated with it; roughly speaking, sampling quality is reflected in typical sizes of error bars, which decrease with better sampling. There does not seem to an unambiguous point where sampling can be considered absolutely converged (although a simulation effectively does not begin to sample equilibrium averages until t_{init} has been surpassed).

It is certainly valuable, nevertheless, to be able to gauge in an absolute sense when sampling is wholly inadequate. Absolute methods that may be useful at detecting extremely poor sampling include casual use of the ergodic measure (86) (i.e., whether or not it approaches zero) and cluster counting (114).

Assessing Dynamical Sampling

The assessment of dynamical methods illustrates the key ideas behind the relative measures of sampling. For this purpose, dynamical is defined as when trajectories consist of configurations with purely sequential correlations—i.e., where a given configuration was produced based solely on the immediately preceding configuration(s). Thus, MD, Langevin dynamics, and simple Markov-chain MC are dynamical methods, but exchange algorithms are not (147). Correlation times or their analogs for MC can then readily be associated with any trajectory that was dynamically generated. Assuming, for the moment, that there is a fundamental

Effective sample size: the number of statistically independent configurations to which a given ensemble of correlated configurations is equivalent

correlation time for overall sampling of the system, t_{corr}^*, then an effective sample size can be calculated from the simple relation

RE: replica exchange

$$N^{eff} \simeq t_{sim}/t_{corr}^*, \qquad 4.$$

where t_{sim} is the total CPU time (or number of cycles) used to generate the trajectory. Thus, sampling quality is relative in that it should increase linearly with t_{sim} (86). In rough terms, the absolute lack of sampling would correspond to $N^{eff} \lesssim 1$; physically, this would indicate that important parts of configuration have been visited only once—e.g., a simulation shorter than $t_{init} + t_{corr}^*$. (It would seem impossible to detect parts of the space never visited.)

Work by Lyman & Zuckerman (73) has suggested that a reasonable overall correlation time t_{corr}^* can be calculated from a dynamical trajectory. Those authors derived their correlation time from the overall distribution in configuration space, estimating the time that must elapse between trajectory frames so that they behave as if statistically independent. The approach has the twin advantages of being based on the full configuration-space distribution (as opposed to isolated observables) and of being blindly and objectively applicable to any dynamical trajectory. Other measures meeting these criteria would also be valuable; note, for instance, the work by Hess (41) using principal components.

Assessing Nondynamical Sampling

If sampling is performed in a nondynamical way, one cannot rely on sequential correlations to assess sampling as in Equation 4. Many of the modern algorithms that attempt to enhance sampling, such as those reviewed below, are not dynamical. Replica exchange (RE) (34, 51, 118) is a typical example: If one is solely interested in the ensemble at a single temperature, a given configuration may be strongly correlated with other configurations distant in the sequence at that temperature and/or be uncorrelated with sequential neighbors. At the same time, RE does generate trajectories that are continuous in configuration space, if not temperature, and it may be possible to analyze these in a dynamical

sense (73) but care will be required (15). Other algorithms, for instance, based on polymer growth procedures (29, 35, 126, 138–140), are explicitly nondynamical.

Two recent papers (38, 141) have argued that nondynamical simulations are best assessed by multiple independent runs. The lack of sequential correlations—but the presence of more complex correlations—in nondynamical ensembles means that the list of configurations cannot be divided into nearly independent segments for blocking-based analysis. Zhang et al. (141) suggest that multiple runs be used to assess variance in the populations of physical states for two related reasons: Such states are defined to be separated from one another by the slowest timescales in a system, and relative populations of states cannot be estimated properly without good sampling within each state. State-population variances, in turn, can be used to estimate N^{eff} based on simple statistical arguments (141); nevertheless, it is important that states be approximated in an automated fashion (see References 14, 91, and 141) to eliminate the possibility for bias.

PURELY ALGORITHMIC EFFORTS TO IMPROVE SAMPLING

How can we beat MD? Despite 30 years of effort, there is no algorithm that is significantly more efficient than MD for the full range of systems of interest. Further, although dozens of different detailed procedures have been suggested, there are a limited number of qualitatively distinct ideas. Some of the strategies that have been proposed are described below.

Replica Exchange and Multiple Temperatures

The most common strategy is to employ elevated temperature. Many variations on this strategy have been proposed; one of the earliest suggestions was for spin systems (27). The approach that has been applied to biomolecules most often is RE (also called parallel
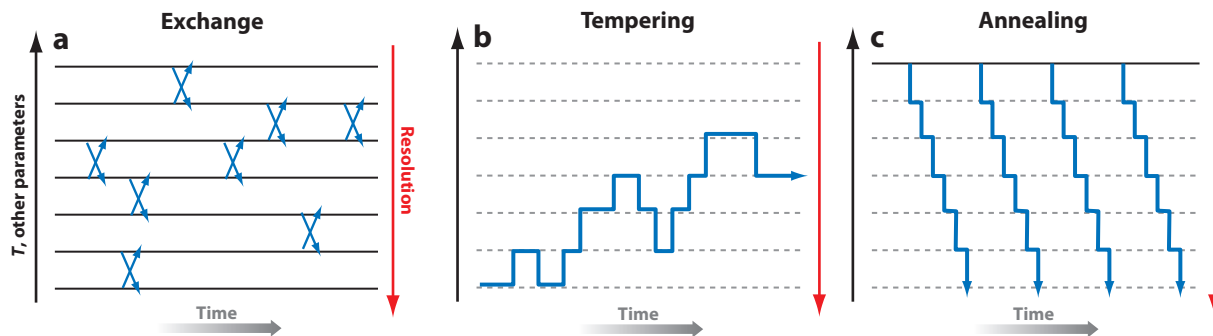
Figure 2

Many ways to use many levels. Sampling algorithms are schematized in the panels: Different horizontal levels represent different simulation conditions such as varying temperature, model parameters, and/or resolution. A given ladder or set of conditions can be used in different formulations: (*a*) exchange, (*b*) tempering, or (*c*) annealing.

tempering), in which parallel simulations wander among a set of fixed temperature values with swaps governed by a Metropolis criterion (34, 51, 118). Many variations and optimizations have been proposed for RE (e.g., 8, 11, 26, 57, 64, 67, 93–95, 99–101, 103, 112, 121). Closely related to RE is simulated tempering, in which a single trajectory wanders among a set of temperatures (75); once again, optimizations have been proposed (e.g., 137) (**Figure 2**).

Less well known, but formally closely related to exchange and tempering, is annealed importance sampling (AIS), in which a high-temperature ensemble is annealed to lower temperatures in a weighted way that preserves canonical sampling (49, 87) (**Figure 2**). AIS has been applied to biomolecules and optimizations have been suggested (72, 74). AIS is nominally a nonequilibrium approach, but in precise analogy to the Jarzynski equality (52), it yields equilibrium ensembles. The J-walking approach also starts from high temperature (28); a good discussion of several related methods is given in Reference 9.

How effective is replica exchange? RE simulation is both popular and controversial. Some authors have noted weaknesses (19, 92, 148), others have described successes (e.g., 48, 104, 102), and some conclusions have been more ambiguous (3, 76, 96, 143). What is the real

story—how much better is RE than ordinary MD?

Examination of the various claims and studies reveals that, in fact, there is little disagreement so long as the particular sampling problem is explicitly accounted for as discussed above. In brief, when observables of interest depend significantly on states that are rapidly accessed at high temperature, then RE and related methods can be efficient (48, 102). A prime example would be estimating the folded fraction near the folding transition. On the other hand, when the goal is to sample an overwhelmingly folded ensemble, there does not appear to be significant evidence of RE's efficiency—so long as the initial condition is a folded configuration (i.e., in terms of the timescales above, if $t_{init} < t_{corr}^*$). This last qualification is important in understanding the data presented in Reference 48.

An intuitive picture underpinning the preceding conclusions is readily gained by revisiting the idea that state populations are estimated by transitions between states in continuous trajectories. Indeed, this same picture is the basis for discrete-state analyses (102, 143), which we can consider along with the schematic landscape of **Figure 1**. Specifically, the average of any observable A can be written as $\langle A \rangle = \sum_i p_i \bar{A}_i$, where i indexes all folded and unfolded substates and $\bar{A}_i$ is the average of A within substate i. If p_i values are significant in both folded and unfolded substates, then

Canonical distribution: the distribution of configurations in which the probability of each is proportional to the Boltzmann factor of the energy at constant temperature, volume, and number of atoms

the elevated temperatures employed in RE simulation can usefully promote transitions to sample unfolded states (102). On the other hand, if $p(\text{unfolded}) \ll p(\text{folded})$, where $p(\text{folded}) = \sum_i^{\text{folded}} p_i$ is the sum of folded probabilities and $p(\text{unfolded}) = 1 - p(\text{folded})$, then computer resources devoted to unfolded substates can detract from estimating folded-state observables; this line of analysis echoes earlier studies (92, 143).

Whether or not RE is superior to MD, another key issue is whether RE can provide sufficient sampling ($N^{\text{eff}} > 10$) for protein-sized systems. It is far from clear that this is the case. RE cannot provide full sampling unless some level of the ladder can be fully sampled (148; see above), and the ability for a simple dynamical trajectory to sample the large configuration space of a fully or partly unfolded state of a protein has not been demonstrated to date.

Energy-based sampling methods. Energy-based schemes must be seen as closely related to temperature-based strategies because the two variables are conjugate to each other in the Boltzmann factor (Equation 1). Yet because any fixed-temperature ensemble can include arbitrary energy values, energy-based schemes must account for this. Multicanonical schemes based on sampling energy values (typically uniformly) have been implemented for biomolecules (40, 85), including with the Wang-Landau device (55, 132). Nevertheless, it should be remembered that energy is not expected to be a good proxy for the configurational coordinates of primary interest (**Figure 1**).

Hamiltonian Exchange and Multiple Models

As schematized in **Figure 2**, the different temperature-based methods sketched above can readily be generalized to variations in force field parameters (117), and even resolution as discussed below. This is because the methods are based on the Boltzmann factor, which is a general form that can apply to different force fields U_λ, each characterized by the set of

parameters $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots\}$. Thus, the fundamental distribution Equation 1 can be written more explicitly as

$$\rho(\mathbf{x}; T, \boldsymbol{\lambda}) \propto \exp[-U_\lambda(\mathbf{x})/k_B T]. \qquad 5.$$

The different schemes for using ladders of different types—based on temperature, force field parameters, or resolution—are depicted in **Figure 2**. One of the first proposals for using multiple models was given in Reference 69.

The simplest way to reduce the roughness of a landscape (cf. Equation 3) is to maintain the functional form of U but change parameters—e.g., coefficients of dihedral terms. Changing parameters in parameter files of common software packages makes this route fairly straightforward. Note that unless force field terms are explicitly removed, the step cost in Equation 3 remains unchanged.

Hamiltonian exchange has been applied in a number of cases, for instance, using models with softened van der Waals' interactions and hence decreased roughness (45). The approach called accelerated MD employs ordinary MD on a smoothed potential energy landscape and requires reweighting to obtain canonical averages (39, 109); it can therefore be considered a two-level Hamiltonian-changing algorithm.

Resolution Exchange and Multiple-Resolution Approaches

Extending the multiple-model ideas a step further leads to the consideration of multiple resolutions. To account for varying resolution in the formulation of Equation 5, some of the λ_i parameters may be considered prefactors that can eliminate detailed interactions when set to zero. Different, formally exact approaches to using multiple resolutions have been proposed (16, 68, 70, 71, 145), primarily echoing ideas in RE and annealing.

Although multiple-resolution methods have not produced dramatic results for canonical sampling of large atomistic systems, they appear to hold unique potential in the context of the sampling costs embodied in Equation 3. In particular, low-resolution models reduce

cost and roughness to such an extent that some coarse models permit good sampling for proteins ($N^{eff} > 10$) with typical resources (77, 135). It is not clear that a similar claim can be made for high-temperature or modified-potential simulations of stably folded proteins.

Multiple Trajectories Without Multiple Levels

There is a qualitatively different group of algorithms that also uses multiple trajectories, but all under the same conditions—i.e., unicanonically (6, 47, 89)—in contrast to RE and related methods. The basic idea of these methods is that when a trajectory reaches a new region of configuration space, it can spawn multiple daughter trajectories, all on the unaltered landscape. The daughter trajectories enable better sampling of regions that otherwise would receive less computer resources, and resources can be directed away from regions that already are well sampled. In the context of Markov-state models, this strategy can be optimized to increase precision in the estimate of any observable (42). Exact canonical sampling can also use a steady-state formulation of the weighted ensemble path sampling method (6; for related approaches see References 98, 125, and 133).

Single-Trajectory Approaches: Dynamics, Monte Carlo, and Variants

It is useful to combine a large group of approaches—including MD, simple MC, and Langevin dynamics (2, 32, 105)—and consider them basic dynamics (147). These methods generate a single trajectory (or Markov chain) in which configurations are strongly sequentially correlated and have no nonsequential correlations. In other words, all these methods should be expected to behave roughly like MD simulation because the motion is intrinsically constrained by the high density and landscape complexity of biomolecular systems. On the one hand, basic dynamical methods generate physically realizable trajectories or nearly so; on the

other hand, they are severely limited by physical timescales as described above. In this section, we explore some basic ideas of dynamical approaches and also some interesting variants.

Are there real dynamics? Unlike many other sampling algorithms, MD also simulates the fundamental classical dynamics of a system. That is, the trajectories produced by MD are also intended to model the time dependence of physical processes—over and above their value for sampling.

Should MD trajectories provide much better depictions of the underlying processes than other basic simulation methods, e.g., Langevin or MC simulation? The answer is not so clear. First, like other methods, MD necessarily uses an approximate force field (31). Second, on the long timescales ultimately of interest, roundoff errors can be expected to lead to significant deviations from the exact trajectories for the given force field.

Somewhat more fundamentally, there appears to be a physical inconsistency in running finite-temperature MD simulations for finite-sized systems. Thermostats of various types are used (32), some stochastic and some deterministic: Is there a correct method? By definition, a finite-temperature system is not isolated but is coupled to a bath by some physical process, presumably molecular collisions. Because the internal degrees of freedom of a bath are not explicitly simulated, again by definition, a bath should be intrinsically stochastic. The author is unaware of a first-principles prescription for modeling the coupling to a thermal bath, although sophisticated thermostats have been proposed (32, 63). For MD simulations with periodic boundary conditions, it is particularly difficult to imagine truly physical coupling to a bath.

A finite-temperature process in a finite system is intrinsically stochastic, rendering questionable the notion that standard MD protocols produce correct trajectories. To consider this point another way, should we really expect inertial effects to be important in dense aqueous and macromolecular media?

Markov chain: a sequence of configurations in which each is generated by a stochastic rule depending on only the immediately prior configuration

The foregoing discussion suggests, certainly, that Langevin simulations should also be considered physical, but what about Monte Carlo? Certainly, one can readily imagine unphysical MC protocols (e.g., attempting trial moves for atoms in the C-terminal domain of a protein ten times as often as for those in the N terminus). Nevertheless, so long as trial moves are small and performed in a spatially uniform way, one can expect MC dynamics to be an approximation to overdamped (noninertial) Langevin simulation (90, 110, 146).

Modifying dynamics to improve sampling. Although changing temperature or model parameters automatically changes the dynamics observed in MD or LD simulation, the dynamics can be modified and maintain canonical sampling on the original landscape, $U(\mathbf{x})$ (84, 144). This possibility defies the simple formulation of the second line of Equation 3, because the landscape roughness remains unchanged while the number of required steps is reduced. The method has been applied to alkane and small-peptide systems, but its generalization to arbitrary systems has not been presented to the author's knowledge. A better known and related approach is the use of multiple time steps, with shorter steps for faster degrees of freedom (119, 122), which can save computer time up to a limit set by resonance phenomena (106).

Another approach to modifying dynamics is to avoid revisiting regions already sampled by means of a memory potential that repels the trajectory from prior configurations (50, 60). This strategy is called metadynamics, or the local-elevation method, and typically employs preselection of important coordinates to be assigned repulsive terms. Although trajectories generated in a history-dependent manner do not satisfy detailed balance, it is possible to correct for the bias introduced and recover canonical sampling (10). This approach could also be considered a potential-of-mean-force method.

The potential surface and dynamics can also be modified by using a strategy of raising basins while keeping barriers intact (39, 97, 129, 149). The accelerated MD approach is a well-known

example, but achieving canonical sampling requires a reweighting procedure (109). Such reweighting is limited by the overlap issues confronting many methods: The sampled distribution must be sufficiently similar to the targeted distribution in Equation 1 so that the data are not overwhelmed by statistical noise (72, 109).

Qualitatively, how can we think about the methods just described? As dynamical methods, they all gather information by transitions among states : More transitions suggest greater statistical quality of the data. However, methods that modify the potential surface may increase the number of transitions, but the overlap between the sampled and targeted distributions is generally expected to decline with more substantial changes to the potential. In this context, the modification of dynamics without perturbing the potential (84, 144) seems particularly intriguing, even though technical challenges in implementation may remain.

Monte Carlo approaches. The term Monte Carlo can be used in many ways, so it useful to first delimit our discussion. Here, we want to focus on single-Markov-chain Monte Carlo, in which a sequence of configurations is generated, each one based on a trial move away from the previous configuration. Typically, such MC simulations of biomolecules have a strong dynamical character because trial moves tend to be small and physically realizable in a small time increment (90, 110, 146), whereas large trial moves would tend to be rejected in any reasonably detailed model. As we shall see, however, less physical moves sometimes can be used—including move sets that do not strictly obey detailed balance but instead conform to the weaker balance condition (79). The various exchange simulations discussed above can be considered MC simulations because exchange is a special kind of trial move and indeed governed by a Metropolis acceptance criterion (147).

A key advantage of single-chain MC simulation is that it can be used with any energy function, whether continuous or not, because forces do not need to be calculated (32). MC simulation is available for use with standard

force fields in some simulation packages (46, 53), but it seems to be most commonly used in connection with simplified models (110, 146). The choice to use MC is usually for convenience: It readily enables the use of simple or discontinuous energy functions; furthermore, MC naturally handles constraints such as fixed bond lengths and/or angles. Substantial effort has gone into developing trial moves useful for biomolecules, with a focus on moves that act fully locally (5, 113, 123, 127, 134). Two studies have reported that MC employing quasi-physical trial moves can be more efficient than MD. (53, 124). MC has also been applied to large implicitly solvated peptides (80). Further, it is widely used in nonstatistical approaches such as docking (62), and it also is used in the Rosetta folding software (18).

For implicitly solvated all-atom peptides, a novel class of trial moves has proven extremely efficient. Ding et al. (20) showed that when libraries of configurations for each amino acid were computed in advance, swap moves with library configurations provided extremely rapid motion in configuration space. Efficiency gains based on measuring N^{eff} suggested the simulations were 100–1,000 times faster than Langevin dynamics. Trial moves involving entire residues apparently were successful because subtle correlations among the atoms were accounted for in the precomputed libraries; by comparison, internal-coordinate trial moves of a similar magnitude (e.g., twisting ϕ and ψ dihedrals) were rarely accepted (20). More detailed discussion of Monte Carlo methods can be found in recent reviews (23, 128) and textbooks (2, 32).

Potential-of-Mean-Force Methods

Many methods are designed to calculate a potential of mean force (PMF) for a specified set of coordinates χ, and such approaches implicitly are sampling methods. After all, the Boltzmann factor of the PMF is defined to be proportional to the probability distribution for the specified coordinates: $\rho(\chi) \propto \exp[-\text{PMF}(\chi)/k_B T]$ (147). If the PMF has been calculated well (with

sufficient sampling along the χ coordinates and orthogonal to them), then the configurations can be suitably reweighted into a canonical ensemble.

Numerous approaches are geared toward calculating a PMF. Some, such as lambda dynamics, explore the full configuration space in a single trajectory (58, 120), and can be categorized with modified dynamics methods, although this is just a semantic issue. Many PMF calculations employ the weighted histogram analysis method (WHAM) (59), but alternatives have been developed (1, 25, 81, 111, 142).

The main advantage that PMF methods hope to attain is faster sampling along the χ coordinates than is possible using brute-force sampling. This perspective allows several observations. First, to aid sampling, the investigator must be able to select all important slow coordinates in advance—otherwise, sampling in directions orthogonal to χ will be impractical. Second, the maximum advantage that can be gained over a brute-force simulation depends on how substantial the barriers are along the χ coordinates; if sampling along a coordinate is slow because many states are separated by only small barriers, then the advantage may be modest. Last, for a PMF calculation to be successful, different values of the χ coordinates must be quantitatively related to each other: Depending on the details of the method, this can occur via numerous transitions (e.g., 25, 58, 120, 142) as given above, or by requiring well-sampled subregions of overlap (59).

Nondynamical Methods

Methods that do not rely on dynamics for sampling use a variety of distinct strategies, but as there are relatively few such efforts, they are grouped together for convenience. Recently, for example, there have been a number of applications to biomolecules of old polymer growth ideas (83, 131), sometimes termed sequential importance sampling (66). Here, monomers (e.g., amino acid) are added one at a time to an ensemble of partially grown configurations while (*a*) keeping track of appropriate statistical

PMF: potential of mean force

weights and/or (*b*) resampling (66). These approaches have been applied to simplified models of proteins and nucleic acids (29, 138–140), to all-atom peptides at high temperature (126), and more recently to all-atom peptides by using presampled libraries of amino acid configurations (78). The intrinsic challenge in polymer growth methods is that configurations important in the full molecule may have low statistical probability in early stages of growth; thus, application to large, detailed systems likely will require biasing toward structural information known for the full molecule (78, 138).

Some methods attempt a semienumerative description of energy basins. For instance, the mining minima method uses a quasi-harmonic procedure to estimate the partition function of each basin, which in turn determines the relative probability of the basin (7, 12). Although such approaches have the disadvantage of requiring an exponentially large number of basins (130), they largely avoid timescale issues associated with dynamical methods because the basins are located with faster search methods.

Another class of approaches has attempted to treat the problem of estimating the free energy of a previously generated canonical ensemble (13, 43, 54, 136). In analogy to PMF methods, knowledge of free energies for independent simulations in separate states (for which no transitions have been observed) enables the states to be combined in a full ensemble: The free energies directly imply the relative probabilities of the states. The ideas behind such methods are fairly technical and beyond the scope of this review.

SPECIAL HARDWARE USE FOR SAMPLING

In several cases, rather impressive improvements in sampling speed have been reported based on using novel hardware—and based on novel uses of ordinary hardware. It is almost always the case that using new hardware also requires algorithmic development, and that component should not be minimized. Little is plug and play in this business.

Parallelization, Special-Purpose CPUs, and Distributed Computing

Perhaps the best known way to exploit hardware is by parallelization. On the one hand, parallelization typically makes sampling less efficient than single-core simulation when measured by sampling per core (or per dollar) due to overhead costs. On the other hand, parallelization allows by far the fastest sampling for a given amount of wall-clock time (for a single system) including record-setting runs (22, 30, 108). Recent examples of parallelization of single-trajectory MD include a 10-μs simulation of a small domain (30); microsecond and longer simulations of membrane proteins on the BlueGene (37) and Desmond (21) platforms; and the longest reported to date, a millisecond simulation of a small globular protein on the Anton machine (24, 108). The Anton simulation reflects parallelization and the use of special-purpose chips, and the chips alone can be inferred to provide a speedup of at least 10 times compared to standard chips (107, 108).

Long simulations have significant value. They allow the community to study selected systems in great detail and to appreciate phenomena that could not otherwise be observed. Of equal importance, long simulations can alert the community to limitations of MD and force fields (31, 36).

There are other parallel strategies, such as exchange simulation (see above) and distributed computing. Distributed computing employs many simultaneous independent simulations or—via repeated rounds of simulation—quasi-parallel computing with minimal communication among simulations. Although distributed computing has been applied primarily to the folding problem (115), recent work has shown the value of multiple short simulations for producing Markov-state models (47, 89). Such models can be used to deduce both nonequilibrium and equilibrium information—thus canonical ensembles if desired. A distributed computing framework can also be used for multilevel simulations such as RE (100) and in principle for other quasi-parallel methods (6).

Use of GPUs and RAM for Sampling

Means other than parallelization are used to exploit hardware. Implicit-solvent simulations of all-atom proteins can be performed hundreds of times faster by combining a GPU and a CPU than with a CPU alone (33). Other work has exploited the RAM available on modern computers: When libraries of amino acid configurations were precalculated, library-based Monte Carlo sampled implicitly solvated peptides hundreds and sometimes more than 1,000 times faster than standard simulation (20). In a similar spirit, tabulation of a scaled form of the generalized Born implicit-solvent model led to significant speed gain for the tabulated part (61).

CONCLUDING PERSPECTIVE

The main goal of this review has been to array various methodologies into qualitative groupings to aid the critical analysis necessary to make progress in equilibrium sampling of biomolecules. Where appropriate, an effort has been made to offer a point of view on essential strengths and weaknesses of various methods. As an example, when considering popular multilevel schemes (see above), users should be confident that some level of the ladder can be well sampled.

Interesting conclusions result from surveying the sampling literature. Except in small systems, purely algorithmic improvements have yet to demonstrably accelerate equilibrium sampling of biomolecules by a significant amount. Hardware-based advances have been more dramatic, however. In fairness, demonstrating the effectiveness of new hardware for MD is much more straightforward than assessing an algorithm.

To aid future progress, developing and using sampling yardsticks should be a key priority for the field. Such measures should probe the configuration-space distribution in an objective, automatic way to measure the effective sample size. Once a small number of standard measures of sampling quality are accepted and used, efforts naturally will focus on approaches that make a significant difference. Currently, there is a proliferation of nuanced modifications of a small number of central ideas, without a good basis for distinguishing among them. Unlike in the history of theoretical physics, where elegance has sometimes served as a guide for truth, the sampling problem cries out for a pragmatic focus on efficiency. After all, it is efficiency that ultimately permits us to address biophysical and biochemical questions with confidence.

In summary, there seems little choice but to express pessimism on the current state of equilibrium sampling of important biomolecular systems. Keys to moving forward would seem to be (*a*) exploiting hardware, (*b*) quantifying sampling to determine which algorithms are more efficient than MD, and (*c*) employing large-resource simulation data to provide benchmarks and guide future efforts.

SUMMARY POINTS

1. The subtleties in defining the equilibrium sampling problem have to do with initial conditions and the targeted number of independent configurations.

2. MD simulations on typical hardware remain several orders of magnitude shorter than known biological timescales.

3. Dozens of algorithmic variants are available for biomolecular simulation, but none has been documented to yield an order of magnitude improvement over standard MD in the range of systems and conditions of primary interest.

4. Algorithms typically have not been assessed by a standard measure, but some such measures are now available.

5. Isolated instances of high algorithmic efficiency (compared to MD) have been reported in small systems and/or special conditions.

6. Novel uses of hardware (CPU, GPU, and RAM) have yielded some of the most dramatic and demonstrable advances.

FUTURE ISSUES

1. Better sampling should reveal much more of the biophysics and biochemistry that motivates biomolecular simulations. We still do not know the scales of typical equilibrium fluctuations in large biomolecules.

2. The degree of sampling quality achieved by simulations must be assessed objectively and quantitatively.

3. Progress in sampling can be expected to come from the combination of novel algorithms and novel hardware use.

4. Although current simulations typically use all-atom models at the expense of poor sampling, it is possible that reduced or hybrid all-atom/coarse-grained models will yield a better overall picture of equilibrium fluctuations.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Abrams JB, Tuckerman ME. 2008. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *J. Phys. Chem. B* 112(49):15742–57
2. Allen MP, Tildesley DJ. 1987. *Computer Simulation of Liquids*. Oxford: Oxford Univ. Press
3. Beck DAC, White GWN, Daggett V. 2007. Exploring the energy landscape of protein folding using replica-exchange and conventional molecular dynamics simulations. *J. Struct. Biol.* 157(3):514–23
4. Berne BJ, Straub JE. 1997. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* 7:181–89
5. Betancourt MR. 2005. Efficient Monte Carlo trial moves for polypeptide simulations. *J. Chem. Phys.* 123(17):174905
6. Bhatt D, Zhang BW, Zuckerman DM. 2010. Steady state via weighted ensemble path sampling. *J. Chem. Phys.* 133:014110
7. Bogdan TV, Wales DJ, Calvo F. 2006. Equilibrium thermodynamics from basin-sampling. *J. Chem. Phys.* 124(4):044102

8. Brenner P, Sweet CR, VonHandorf D, Izaguirre JA. 2007. Accelerating the replica exchange method through an efficient all-pairs exchange. *J. Chem. Phys.* 126(7):074103

9. Brown S, Head-Gordon T. 2002. Cool walking: a new Markov chain Monte Carlo sampling method. *J. Comp. Chem.* 24:68–76

10. Bussi G, Laio A, Parrinello M. 2006. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* 96(9):090601

11. Calvo F. 2005. All-exchanges parallel tempering. *J. Chem. Phys.* 123(12):124106

12. Chang CE, Chen W, Gilson MK. 2007. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. USA* 104(5):1534–39

13. Cheluvaraja S, Meirovitch H. 2004. Simulation method for calculating the entropy and free energy peptides and proteins. *Proc. Natl. Acad. Sci. USA* 101:9241–46

14. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. 2007. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126(15):155101–17

15. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA. 2007. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.* 3(1):26–41

16. Christen M, van Gunsteren WF. 2006. Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. *J. Chem. Phys.* 124(15):154106

17. Christen M, van Gunsteren WF. 2008. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review. *J. Comput. Chem.* 29(2):157–66

18. Das R, Baker D. 2008. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* 77:363–82

19. Denschlag R, Lingenheil M, Tavan P. 2008. Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria. *Chem. Phys. Lett.* 458:244–48

20. Ding Y, Mamonov AB, Zuckerman DM. 2010. Efficient equilibrium sampling of all-atom peptides using library-based Monte Carlo. *J. Phys. Chem. B* 114(17):5870–77

21. Dror RO, Arlow DH, Borhani DW, Jensen MÃ, Piana S, Shaw DE. 2009. Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. USA* 106(12):4689–94

22. Duan Y, Kollman PA. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282(5389):740–44

23. Earl DJ, Deem MW. 2008. Monte Carlo simulations. *Methods Mol. Biol.* 443:25–36

24. Ensign DL, Kasson PM, Pande VS. 2007. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J. Mol. Biol.* 374(3):806–16

25. Fasnacht M, Swendsen RH, Rosenberg JM. 2004. Adaptive integration method for Monte Carlo simulations. *Phys. Rev. E* 69:056704

26. Fenwick MK, Escobedo FA. 2003. Expanded ensemble and replica exchange methods for simulation of protein-like systems. *J. Chem. Phys.* 119(22):11998–2010

27. Ferrenberg AM, Swendsen RH. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–38

28. Frantz DD, Freeman DL, Doll JD. 1990. Reducing quasi-ergodic behavior in Monte Carlo simulation by J-walking: applications to atomic clusters. *J. Chem. Phys.* 93:2769–84

29. Frauenkron H, Bastolla U, Gerstner E, Grassberger P, Nadler W. 1998. New Monte Carlo algorithm for protein folding. *Phys. Rev. Lett.* 80(14):3149–52

30. Freddolino PL, Liu F, Gruebele M, Schulten K. 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* 94(10):L75–77

31. Freddolino PL, Park S, Roux B, Schulten K. 2009. Force field bias in protein folding simulations. *Biophys. J.* 96(9):3772–80

32. Frenkel D, Smit B. 1996. *Understanding Molecular Simulation*. San Diego, CA: Academic

33. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, et al. 2009. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30:864–72

34. Geyer CJ. 1991. Markov chain Monte Carlo maximum likelihood. In *Comput. Sci. Stat.: Proc. 23rd Symp. Interface*, ed. EM Keramidas, pp. 156–63. Fairfax Station: Interface Found.

33. Describes GPU-based simulation of implicitly solvated proteins which are hundreds of times faster than CPU simulation.

35. Grassberger P. 1997. Pruned-enriched Rosenbluth method: simulations of theta polymers of chain length up to 1,000,000. *Phys. Rev. E* 56(3):3682–93

36. Grossfield A, Feller SE, Pitman MC. 2007. Convergence of molecular dynamics simulations of membrane proteins. *Proteins* 67(1):31–40

37. Grossfield A, Pitman MC, Feller SE, Soubias O, Gawrisch K. 2008. Internal hydration increases during activation of the G-protein-coupled receptor rhodopsin. *J. Mol. Biol.* 381(2):478–86

38. Grossfield A, Zuckerman DM. 2009. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 5:23–48

39. Hamelberg D, Mongan J, McCammon JA. 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120(24):11919–29

40. Hansmann UH, Okamoto Y. 1996. Monte Carlo simulations in generalized ensemble: multicanonical algorithm versus simulated tempering. *Phys. Rev. E* 54(5):5863–65

41. Hess B. 2002. Convergence of sampling in protein simulations. *Phys. Rev. E* 65(3):031910

42. Hinrichs NS, Pande VS. 2007. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.* 126(24):244101

43. Hnizdo V, Tan J, Killian BJ, Gilson MK. 2008. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J. Comput. Chem.* 29(10):1605–14

44. Howard J. 2001. *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer

45. Hritz J, Oostenbrink C. 2008. Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* 128(14):144121

46. Hu J, Ma A, Dinner AR. 2006. Monte Carlo simulations of biomolecules: the MC module in CHARMM. *J. Comput. Chem.* 27(2):203–16

47. Huang X, Bowman GR, Bacallado S, Pande VS. 2009. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl. Acad. Sci. USA* 106(47):19765–69

48. Huang X, Bowman GR, Pande VS. 2008. Convergence of folding free energy landscapes via application of enhanced sampling methods in a distributed computing environment. *J. Chem. Phys.* 128(20):205106

49. Huber GA, McCammon JA. 1997. Weighted-ensemble simulated annealing: faster optimization on hierarchical energy surfaces. *Phys. Rev. E* 55(4):4822

50. Huber T, Torda AE, van Gunsteren WF. 1994. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput. Aided Mol. Des.* 8(6):695–708

51. Hukushima K, Nemoto K. 1996. Exchange Monte Carlo method and application to spin glass simulation. *J. Phys. Soc. Jpn.* 65:1604–8

52. Jarzynski C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–93

53. Jorgensen WL, Tirado-Rives J. 1996. Monte Carlo vs molecular dynamics for conformational sampling. *J. Phys. Chem.* 100:14508–13

54. Killian BJ, Kravitz JY, Gilson MK. 2007. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* 127(2):024107

55. Kim J, Straub JE, Keyes T. 2006. Statistical-temperature Monte Carlo and molecular dynamics algorithms. *Phys. Rev. Lett.* 97(5):050601

56. Knegtel RMA, Kuntz ID, Oshiro CM. 1997. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266:424–40

57. Kofke DA. 2002. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* 117(15):6911–14

58. Kong X, Brooks CM. 1996. Lambda-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* 105:2414–23

59. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.* 13:1011–21

60. Laio A, Parrinello M. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* 99(20):12562–66

61. Larsson P, Lindahl E. 2010. A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling. *J. Comput. Chem.* 31(14):2593–600

62. Leach AR. 2001. *Molecular Modelling: Principles and Applications*. Upper Saddle River, NJ: Prentice Hall

63. Leimkuhler BJ, Sweet CR. 2004. The canonical ensemble via symplectic integrators using nose and nose-poincare chains. *J. Chem. Phys.* 121(1):108–16

64. Li H, Li G, Berg BA, Yang W. 2006. Finite reservoir replica exchange to enhance canonical sampling in rugged energy surfaces. *J. Chem. Phys.* 125(14):144902

65. Lin J-H, Perryman AL, Schames JR, McCammon JA. 2002. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* 124(20):5632–33

66. Liu JS. 2002. *Monte Carlo Strategies in Scientific Computing*. New York: Springer

67. Liu P, Kim B, Friesner RA, Berne BJ. 2005. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* 102:13749–54

68. Liu P, Shi Q, Lyman E, Voth GA. 2008. Reconstructing atomistic detail for coarse-grained models with resolution exchange. *J. Chem. Phys.* 129(11):114103

69. Liu Z, Berne BJ. 1993. Method for accelerating chain folding and mixing. *J. Chem. Phys.* 99(8):6071–77

70. Lyman E, Ytreberg FM, Zuckerman DM. 2006. Resolution exchange simulation. *Phys. Rev. Lett.* 96:028105

71. Lyman E, Zuckerman DM. 2006. Resolution exchange simulation with incremental coarsening. *J. Chem. Theory Comput.* 2:656–66

72. Lyman E, Zuckerman DM. 2007. Annealed importance sampling of peptides. *J. Chem. Phys.* 127(6):065101–6

73. Lyman E, Zuckerman DM. 2007. On the structural convergence of biomolecular simulations by determination of the effective sample size. *J. Phys. Chem. B* 111(44):12876–82

74. Lyman E, Zuckerman DM. 2009. Resampling improves the efficiency of a "fast-switch" equilibrium sampling protocol. *J. Chem. Phys.* 130(8):081102

75. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN. 1992. New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. *J. Chem. Phys.* 96(3):1776–83

76. Machta J. 2009. Strengths and weaknesses of parallel tempering. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* 80(5 Pt. 2):056706

77. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. 2009. General library-based Monte Carlo technique enables equilibrium sampling of semi-atomistic protein models. *J. Phys. Chem. B* 113(31):10891–904

78. Mamonov AB, Zhang X, Zuckerman D. 2011. Rapid sampling of all-atom peptides using a library-based polymer-growth approach. *J. Comput. Chem.* 32:396–405

79. Manousiouthakis VI, Deem MW. 1999. Strict detailed balance is unnecessary in Monte Carlo simulation. *J. Chem. Phys.* 110(6):2753–56

80. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* 107(18):8183–88

81. Maragliano L, Vanden-Eijnden E. 2008. Single-sweep methods for free energy calculations. *J. Chem. Phys.* 128(18):184110

82. McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature* 267(5612):585–90

83. Meirovitch H. 1982. A new method for simulation of real chains. Scanning future steps. *J. Phys. A* 15:L735–40

84. Minary P, Tuckerman ME, Martyna GJ. 2004. Long time molecular dynamics for enhanced conformational sampling in biomolecular systems. *Phys. Rev. Lett.* 93(15):150201

85. Mitsutake A, Sugita Y, Okamoto Y. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 60(2):96–123

86. Mountain RD, Thirumalai D. 1989. Measures of effective ergodic convergence in liquids. *J. Phys. Chem.* 93(19):6975–79

87. Neal RM. 2001. Annealed importance sampling. *Stat. Comput.* 11:125–39

88. Neirotti JP, Freeman DL, Doll JD. 2000. Approach to ergodicity in Monte Carlo simulations. *Phys. Rev. E* 62(5):7445–61

89. Noe F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. 2009. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA* 106(45):19011–16

90. Nowak U, Chantrell RW, Kennedy EC. 2000. Monte Carlo simulation with time step quantification in terms of Langevin dynamics. *Phys. Rev. Lett.* 84(1):163–66

77. Introduces a memory-intensive approach to simulation of all-atom, coarse-grained, and hybrid models.

91. Noé F, Horenko I, Schütte C, Smith JC. 2007. Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.* 126(15):155102

92. Nymeyer H. 2008. How efficient is replica exchange molecular dynamics? An analytic approach. *J. Chem. Theory Comput.* 4(4):626–36

93. Okur A, Roe DR, Cui G, Hornak V, Simmerling C. 2007. Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J. Chem. Theory Comput.* 3(2):557–68

94. Opps SB, Schofield J. 2001. Extended state-space Monte Carlo methods. *Phys. Rev. E* 63(5 Pt. 2):056701

95. Paschek D, Garcia AE. 2004. Reversible temperature and pressure denaturation of a protein fragment: a replica-exchange molecular dynamics simulation study. *Phys. Rev. Lett.* 93:238105

96. Periole X, Mark AE. 2007. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.* 126(1):014903

97. Rahman JA, Tully JC. 2002. Puddle-skimming: an efficient sampling of multidimensional configuration space. *J. Chem. Phys.* 116(20):8750–60

98. Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M. 2006. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B* 110(8):3533–39

99. Rathore N, Chopra M, de Pablo JJ. 2005. Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* 122(2):024111

100. Rhee YM, Pande VS. 2003. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* 84(2 Pt. 1):775–86

101. Rick SW. 2007. Replica exchange with dynamical scaling. *J. Chem. Phys.* 126(5):054102

102. Rosta E, Hummer G. 2009. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* 131(16):165102

103. Ruscio JZ, Fawzi NL, Head-Gordon T. 2010. How hot? Systematic convergence of the replica exchange method using multiple reservoirs. *J. Comput. Chem.* 31(3):620–27

104. Sanbonmatsu KY, García AE. 2002. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins* 46(2):225–34

105. Schlick T. 2002. *Molecular Modeling and Simulation*. New York: Springer

106. Schlick T, Mandziuk M, Skeel RD, Srinivas K. 1998. Nonlinear resonance artifacts in molecular dynamics simulations. *J. Comput. Phys.* 140(1):1–29

107. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, et al. 2007. Anton: a special-purpose machine for molecular dynamics simulation. In *Proc. 34th Annu. Int. Symp. Comput. Archit.*, *San Diego, CA, (ISCA '07)*, pp. 1–12. New York: Assoc. Comput. Mach.

108. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, et al. 2009. Millisecond-scale molecular dynamics simulations on Anton. In *Proc. Conf. High Perform. Comput. Netw., Storage Anal., (SC '09)*, pp. 1–11. New York: Assoc. Comput. Mach.

109. Shen T, Hamelberg D. 2008. A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.* 129(3):034103

110. Shimada J, Kussell EL, Shakhnovich EI. 2001. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *J. Mol. Biol.* 308(1):79–95

111. Shirts MR, Chodera JD. 2008. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129(12):124105

112. Sindhikara D, Meng Y, Roitberg AE. 2008. Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.* 128(2):024103

113. Smith CA, Kortemme T. 2008. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* 380(4):742–56

114. Smith LJ, Daura X, van Gunsteren WF. 2002. Assessing equilibration and convergence in biomolecular simulations. *Proteins* 48:487–96

115. Snow CD, Nguyen H, Pande VS, Gruebele M. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* 420(6911):102–6

116. Straub JE, Thirumalai D. 1993. Exploring the energy landscape in proteins. *Proc. Natl. Acad. Sci. USA* 90(3):809–13

108. A millisecond-scale simulation of an explicitly solvated protein, performed with special hardware and parallelization algorithms.

117. Sugita Y, Kitao A, Okamoto Y. 2000. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113:6042–51

118. Swendsen RH, Wang J-S. 1986. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* 57:2607–9

119. Teleman O, Jonsson B. 1986. Vectorizing a general purpose molecular dynamics simulation program. *J. Comput. Chem.* 7(1):58–66

120. Tidor B. 1993. Simulated annealing on free energy surfaces by a combined molecular dynamics and Monte Carlo approach. *J. Phys. Chem.* 97:1069–73

121. Trebst S, Troyer M, Hansmann UHE. 2006. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* 124(17):174903

122. Tuckerman M, Berne BJ, Martyna GJ. 1992. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 97(3):1990–2001

123. Ulmschneider JP, Jorgensen WL. 2003. Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *J. Chem. Phys.* 118(9):4261–71

124. Ulmschneider JP, Ulmschneider MB, Nola AD. 2006. Monte Carlo vs molecular dynamics for all-atom polypeptide folding simulations. *J. Phys. Chem. B* 110(33):16733–42

125. Valeriani C, Allen RJ, Morelli MJ, Frenkel D, ten Wolde PR. 2007. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. *J. Chem. Phys.* 127(11):114109

126. Velikson B, Garel T, Niel JC, Orland H, Smith JC. 1992. Conformational distribution of heptaalanine: analysis using a new Monte Carlo chain growth method. *J. Comput. Chem.* 13(10):1216–33

127. Vendruscolo M. 1997. Modified configurational bias Monte Carlo method for simulation of polymer systems. *J. Chem. Phys.* 106(7):2970–76

128. Vitalis A, Pappu RV. 2009. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* 5:49–76

129. Voter AF. 1997. Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* 78(20):3908–11

130. Wales D. 2004. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses.* Cambridge, UK: Cambridge Univ. Press

131. Wall FT, Erpenbeck JJ. 1959. New method for the statistical computation of polymer dimensions. *J. Chem. Phys.* 30(3):634–37

132. Wang F, Landau DP. 2001. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86(10):2050–53

133. Warmflash A, Bhimalapuram P, Dinner AR. 2007. Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.* 127(15):154112–18

134. Wu MG, Deem MW. 1999. Analytical rebridging Monte Carlo: application to cis/trans isomerization in proline-containing, cyclic peptides. *J. Chem. Phys.* 111(14):6625–32

135. Ytreberg FM, Aroutiounian SK, Zuckerman DM. 2007. Demonstrated convergence of the equilibrium ensemble for a fast united-residue protein model. *J. Chem. Theory Comput.* 3(5):1860–66

136. Ytreberg FM, Zuckerman DM. 2008. A black-box re-weighting analysis can correct flawed simulation data. *Proc. Natl. Acad. Sci. USA* 105(23):7982–87

137. Zhang C, Ma J. 2010. Enhanced sampling and applications in protein folding in explicit solvent. *J. Chem. Phys.* 132(24):244101

138. Zhang J, Lin M, Chen R, Liang J, Liu JS. 2007. Monte Carlo sampling of near-native structures of proteins with applications. *Proteins* 66(1):61–68

139. Zhang J, Lin M, Chen R, Wang W, Liang J. 2008. Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J. Chem. Phys.* 128(12):125107

140. Zhang JL, Liu JS. 2002. A new sequential importance sampling method and its application to the two-dimensional hydrophobic–hydrophilic model. *J. Chem. Phys.* 117(7):3492–98

141. Zhang X, Bhatt D, Zuckerman DM. 2010. Automated sampling assessment for molecular simulations using the effective sample size. *J. Chem. Theory Comput.* 6:3048–57

142. Zheng L, Chen M, Yang W. 2008. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. USA* 105(51):20227–32

118. Introduces replica exchange simulation.

133. Introduces a method for sampling steady states (of which equilibrium is one).

141. Provides an automated assessment of sampling quality for the effective sample size based on the configuration-space distribution.

143. Describes kinetic analysis of replica exchange explaining strengths and weaknesses.

143. **Zheng W, Andrec M, Gallicchio E, Levy RM. 2007. Simulating replica exchange simulations of protein folding with a kinetic network model.** *Proc. Natl. Acad. Sci. USA* **104(39):15340–45**

144. **Zhu Z, Tuckerman ME, Samuelson SO, Martyna GJ. 2002. Using novel variable transformations to enhance conformational sampling in molecular dynamics.** *Phys. Rev. Lett.* **88(10):100201**

144. Describes a method for accelerating dynamics in the canonical ensemble.

145. Zuckerman D. 2009. Principles and practicalities of canonical mixed-resolution sampling of biomolecules. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. GA Voth, pp. 171–84. Boca Raton, FL: Taylor & Francis

146. Zuckerman DM. 2004. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J. Phys. Chem. B* 108:5127–37

147. Zuckerman DM. 2010. *Statistical Physics of Biomolecules: An Introduction*. Boca Raton, FL: CRC Press

148. Zuckerman DM, Lyman E. 2006. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* 2(4):1200–2

149. Zuckerman DM, Woolf TB. 1999. Dynamic reaction paths and rates through importance-sampled stochastic dynamics. *J. Chem. Phys.* 111:9475–84

Contents

Index

Errata

An online log of corrections to *Annual Review of Biophysics* articles may be found at
http://biophys.annualreviews.org/errata.shtml