Appeared in: “Third International Workshop for Methods for Macromolecular Modeling Conference Proceedings”, H.H. Gan and T. Schlick, eds., Spinger, New York (2002).

On the Potential of Monte Carlo Methods for Simulating Macromolecular Assemblies

Mihaly Mezei

Department of Physiology and Biophysics, Mount Sinai School of Medicine, NYU, New York, NY 10029, USA.

Abstract. A wide variety of Monte Carlo techniques are described to argue that the methodology has a large untapped potential to solve sampling problems for complex systems.

1. Introduction

Computer simulations of assemblies of atoms and molecules [1,2] use almost exclusively one of two well known techniques: Monte Carlo (MC), implemented by variants of the Metropolis method [3] or molecular dynamics (MD), that involves the integration of Newton’s second law of motion [4]. While in the pioneering years of simulations the MC technique was prevalent, simulation work on biomolecular systems in the last decade was dominated by the MD technique. The principal aim of this paper is to make a case for putting more work in developing MC methodologies as there are several features of MC that make it well suited for the resolution of a number of sampling problems.

The power of MD lies in the fact that it is driven by Newton’s laws of motion — a governing principle that is guaranteed to work. Also, explicit introduction of time allows the simple calculation of time-dependent properties. However, using the actual time evolution for the sampling of the configuration space also imposes a limitation on the sampling rate. Also, being governed by a physical principle, the trajectory in the configuration space also has to be physical — in general this means that, first of all, the path has to be essentially continuous.

MC, on the other hand, is free from the constraint of having to follow a physical path, thus it can take shortcuts and jumps, thereby (potentially) significantly improve the sampling efficiency. This potential, however, is balanced by the facts that a) it is not at all easy to find the shortcuts; b) time-dependent properties can only indirectly modeled; c) MC methods that change only a small part of the system each step (as is true for most realizations) are inefficient to use with some of the more complex Hamiltonians: force fields where the energy of the system includes an induction term [5,6] or with the so called Particle-Mesh Ewald (PME) treatment [7] of long-range electrostatics. Note, however, that the formalism introduced by Sperb [8] for the treatment of long-range electrostatics is pairwise additive and thus it may be an efficient alternative to PME.

2. General formalism for the MC method

The MC technique used for simulation of atomic and molecular assemblies, usually referred to as the Metropolis method [3], is based on the construction of a Markov chain whose limiting distribution π is the Boltzmann distribution in the ensemble under consideration. This requires the construction of a transition matrix such that

$$\pi = \pi \mathbf{P} \tag{1}$$

A general form of the elements of $\mathbf{P}$ was shown by Hastings [9] to be

$$p_{ij} = q_{ij} \frac{s_{ij}}{1 + (\pi_i q_{ij})/(\pi_j q_{ji})} \tag{2}$$

where the p_{mn}'s satisfy the reversibility condition

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{3}$$

with

$$p_{ii} = 1 - \sum_{i \neq j} p_{ij} \ , \tag{4}$$

the q_{ij}'s are the elements of the transition matrix of an arbitrary irreducible Markov chain, usually referred to as the *a priori* transition probabilities, on the same states π is defined and s_{ij} is a symmetric function of i and j satisfying

$$0 \leq \frac{s_{ij}}{1 + (\pi_i q_{ij})/(\pi_j q_{ji})} \leq 1 \ . \tag{5}$$

This formulation highlights the many freedoms the MC methodology allows. The original Metropolis method is recovered with the choice of

$$q_{ij} = q_{ji},$$

$$s_{ij} = 1 + (\pi_i q_{ij})/(\pi_j q_{ji}) \quad ((\pi_i q_{ij})/(\pi_j q_{ji}) \geq 1) \tag{6}$$

$$s_{ij} = 1 + (\pi_j q_{ji})/(\pi_i q_{ij}) \quad ((\pi_j q_{ji})/(\pi_i q_{ij}) \leq 1)$$

while the choice

$$q_{ij} = q_{ji} \qquad \text{and} \qquad s_{ij} = 1 \tag{7}$$

results in the prescription suggested by Barker [10], frequently uses in the simulation of spin systems. Translated into words, in the Metropolis method moves with $\pi_j \leq \pi_i$ are always accepted and moves with $\pi_j > \pi_i$ are accepted with probability π_j/π_i while in the rarely used Barker method each move is accepted with probability $\pi_j/(\pi_i + \pi_j)$. Peskun has shown [11] that using the Metropolis method results in more precise estimates of expectation values averaged over the chain than using the Barker method. A so far unexplored possibility exists, however, that the Barker method may be better suited for problems requiring the crossing of barriers since it does not force a 'downward' move any time a trial state with lower π is selected. When the Metropolis prescription is combined with an asymmetric q_{ij}, the acceptance probability is given as [1]

$$\min\{1, (\pi_i q_{ij})/(\pi_j q_{ji})\} \tag{8}$$

The Hastings formalism was recently further generalized by Liu, Liang and Wong [12] to provide a framework to methods that compare several trys before actually selecting a new move.

The choice of q_{ij}'s offer a rich variety of possibilities. In practical terms, q_{ij}'s control the trial changes in the system. In particular, they control a) the subset of the system that will be changed (*e.g.,* the molecule to displace or the torsion angle to change); b) the size of the change (*e.g.,* the range of coordinate values from which the trial state is chosen randomly); and c) the distribution from which the trial change is going to be selected (*e.g.,* uniform or something more 'informed'). Choices with $q_{ij} \neq q_{ji}$ are usually referred to as *biased sampling.* Note, that biased sampling is different from non-Boltzmann (often referred to as 'umbrella') sampling [13,14]: in the latter the Markov chain is constructed based on a modified distribution π', and the Boltzmann averages corresponding to π are recovered using a modified formalism. Umbrella sampling was developed in the context of free energy simulations, but the concept is of more general scope (see Sec. 3.6). Also, while biased sampling is exclusive to MC non-Boltzmann sampling can be performed with both MD and MC.

Applying the Markov chain approach to continuous systems introduces a subtle issue. The usual argument states that in digital computers all states are actually discrete thus the formalism derived based on discrete states carries over. This argument, however, ignores the fact that for continuous systems nonzero Boltzmann probabilities can only result in a finite volume, *i.e.,* instead of stating $P(\mathbf{X}^N) = \exp(-E(\mathbf{X}^N)/kT)$ one should state $P(\mathbf{x}^N \in [\mathbf{X}^N, \mathbf{X}^N + d\mathbf{X}^N]) = \exp(-E(\mathbf{X}^N)/kT)d\mathbf{X}^N$. While in most applications $d\mathbf{X}^N$ is constant and thus cancels when the ratio of probabilities is formed, for attempted changes where there is a correlation among the degrees of freedom varied, $d\mathbf{X}^N$ may also change. This requires the inclusion of the Jacobian associated with such change — an example for this is discussed in Sec. 3.4.

In addition to the selection of the appropriate transition probabilities q_{ij} and the function s_{ij} the correctness of a MC technique is ensured only if the Markov chain generated by it is *ergodic*, *i.e.,* each state j can be reached from each state i in a finite number of steps with non-zero probability. Lack of ergodicity can arise, *e.g.,* if the scope of changes allowed by q_{ij} is too limited or if the order of changes induces a cycle in the states sampled. On the other hand, MD is considered inherently ergodic. While, in general, true ergodicity is rarely a problem, the so-called quasi-ergodicity frequently is, affecting both MD and MC methods. In practical terms, it means that it is often difficult to ensure adequate sampling of the relevant parts of the configuration space within reasonable computing time.

To make the case for the possibilities inherent in the MC methodology, the rest of this paper will present a collection of wide ranging realizations and analyze the issues involved in their application. The emphasis will be on methods related to modeling continuous systems — simulations of discrete models (*e.g.,* lattice models) have always relied heavily on the MC methodology. The selection is *not* meant to be anywhere near comprehensive — that would be beyond the scope of this paper as even the 1984 review of Levesque et al. [15] already contained 467 references. Other resources include a 1999 volume of Advances in Chemical

Physics [16], an archive of papers on Markov Chain Monte Carlo methods [17] and a web site dedicated to Molecular Monte Carlo [18].

3. The selection of the *a priori* transition matrix q_{ij}

The design of a MC method largely consists of the choice of q_{ij}'s. The governing principle of a good design is to make large changes in the system in each step while keeping the corresponding $(\pi_i q_{ij})/(\pi_j q_{ji})$ ratio away from zero, so that the acceptance of the trial change should not become hopeless. This section present a systematic analysis of the choices involved in selecting q_{ij}.

3.1. Selection of the subset of the system to be changed at each step.

The idea of changing only a subset of the system at each step follows from the nature of most potential functions defining the Boltzmann distribution: due to the so-called hard core of atoms (the result of exchange repulsion) the energy can increase very steeply as a function of atomic coordinates. As a result, even small changes in the coordinates of several atoms are very likely result in an overlap between at least one pair, giving rise to a steep increase in the energy, with the concomitant steep drop in π_j, resulting in near-certain rejection of the trial. Thus, as long as the cost of energy calculation is (approximately) additive over the subsets considered, it is generally advantageous to select as small a subset to change in each step as possible, since this way the effect of the occasional sharp increase due to the overlap of even one pair of atoms will not cause the near certain rejection of the rest of the subsystem changes.

There is, however, an important caveat. For more complex systems, judiciously chosen correlated changes of several degrees of freedom can result in smaller change in the Boltzmann probability than any comparable change in the individual degrees of freedom considered. Correlated changes, however, make the behavior of q_{ij}'s more complex. One of the major challenge in the development of efficient MC techniques is the design of such correlated changes.

Once the size of the subsystems to change is settled there is the choice of distribution from which the subsets are selected from change. This choice sets the relative frequencies of various types of moves (*e.g.,* relative frequencies of particle displacement and volume change in the (T, P, N) ensemble) when there are more than one types and the order of choice or distribution to sample from for the selection of the individual subsystems within each type of change. In most applications the distribution is chosen to be uniform and the selection is either generated randomly or in a cycle. A combination of these two approaches uses a random permutation within each cycle [19]. There can be significant benefits, however, from using non-uniform distributions to select subsystems to change. The best known such technique is the preferential sampling of Owicki [20] where solvents near the solute are moved more frequently than solvents far away. For complex systems, preferential sampling of any degree of freedom can be implemented based on proximity to regions of interest, *e.g.,* an active site. For simulations in the grand-canonical ensemble (*vide infra*) the selection of particles to be removed can also be chosen to occur more frequently near regions of interest (with the concomitant increase in the frequency of insertion attempts in that region) [21]. This same principle was applied in Ref. 22 where torsion angles near the end of the chain were changed less frequently than the ones farther from the chain's end.

3.2. Selection of the size of the change

Initial MC work simulating largely homogeneous systems fixed the range within which each degree of freedom can change and selected the actual values from the uniform distribution. However, when the system has significant inhomogeneity (*e.g.,* the density or the distribution of energy is far from uniform), a single range has to be selected in such a way as to make possible accepted moves even in the most inhomogeneous region (*i.e.,* the stepsize has to be small) and this will significantly lower the efficiency of move attempts elsewhere.

Goldman has suggested to modulate the stepsize as a function of the energy of the moved molecule [23,24]: the lower the energy, the smaller are the moves that are likely to be accepted. This requires the use of Eq. (8), with the additional twist that a trial change resulting in a lower energy may have to be rejected because the initial state can be outside the range of the trial state. This technique was also found to accelerate barrier crossings [25]. The same approach has also been applied to simulation of systems with density inhomogeneity [26].

A recent application extended the idea to sampling in the torsion-angle space [22]. Here the torsion angle stepsize range was modulated by the distance of the atom that is farthest from the rotation axis (among the atoms affected by the torsion).

3.3. Biased selection of the change

While most MC applications select the trial changes from uniform distribution, improved sampling can be achieved if additional information about the system can be incorporated into the selection process. The prime example of such technique is the force-biased sampling of displacement of Rao, Pangali, and Berne [27]. The change in each coordinate i is sampled from the distribution

$$P(x_i) = c_i^N \exp(-\lambda F_i \Delta x_i / kT) \tag{9}$$

where k is the Boltzmann constant, T is the absolute temperature the simulation is run at, F_i is the force component along the coordinate x_i, Δx_i is the range of changes in x_i c_i^N is the normalization factor calculated to make $P(x_i)$ integrate out to unity and λ is a constant, usually taken to be 1/2. Analogous expression is used to sample the orientational change using torque components instead of the force components. This technique was found to improve the sampling efficiency for the simulation of liquid water by a factor of 2 to 4 [27,28] at the cost of only about 30% extra computing. A different variant introduced by Cao end Berne, called “anti-force bias”, applies this scheme only in the convex regions of the potential surface and reverses the direction of the bias in other places [25]. The purpose of this change is to facilitate barrier crossing.

The idea of force-biasing has also been extended to the sampling of volume changes in the (T, P, N) ensemble [29]. Here the role of force is played by the virial sum thus the added computational burden is minimal if the displacement attempts use force biasing, since in that case the force components (the computationally expensive contribution to the virial sum) are already present.

Force biasing also provided an example to the always present danger that a perceived ‘short-cut’ in the configuration space leads instead to ‘getting lost’: a simulation of an aqueous system containing (monatomic) ions with force-biasing lead to a situation where some waters were never moved during a reasonably long stretch. It turned out to be the result of the water always trying to move toward the ion, bumping into its repulsive wall. This conundrum can be resolved by making λ a function of the distance from the ionic species (in case of more than one ion, of the distance from the nearest one) [30].

It is also possible to sample changes from a distribution that has been established from prior knowledge of the system. For example, Hardy and Pastor [31] used the dominant part of the energy function to generate an approximate Boltzmann factor for the sampling of a lipid molecule in a mean field. Guarnieri and Weinstein [32] used successfully the idea of 'conformational memories', *i.e.,* a distribution of feasible conformations, determined from several independent high-temperature runs, to efficiently sample the conformations of a decapeptide.

3.4. Correlated changes

The major difficulty in applying the MC methodology to macromolecular systems lies in the fact that, due to the very large force constant associated with a chemical bond, the energy of the system varies very steeply when a single atom's coordinate is changed, limiting the practical changes to ranges too small to be of practical use. Thus, successful MC methods have to move several atoms, leading to the problem of designing such moves with well defined q_{ij}'s.

A rather successful realization of this strategy is the idea of performing simulations in the torsion angle space, as realized early in the ECEPP suite of programs and force field from the Scheraga Laboratory [33 and references therein]. As this approach avoids changing bond lengths and bond angles (the contributors to the steepest part of the potential) it is very efficient for small molecules. In fact, Jorgensen and Tirado-Rives have shown [34] that the conformational space of liquid hexane can be sampled significantly better with MC (even when bond-angle changes are also allowed) than with MD.

Once the size of the molecule is increased, the efficiency of a move attempt changing a single torsion angle drops precipitously, since the change in atomic coordinates is roughly a linear function of the distance of the atom from the torsion bond. Thus, macromolecular applications require additional ideas.

The first such idea has been introduced by Noguti and Go [35] with their Scaled Collective Variables (SCV) method. The SCV η_i are defined as

$$\eta_i = \lambda_i^{1/2} \sum_i u_{ij} \phi_j \tag{10}$$

where u_{ij} and λ_i are the eigenvectors and eigenvalues of the second derivative matrix of the energy (w.r.t. the torsion angles ϕ_i). Unlike the local moves discussed below, the volume element involved in the coordinate transformation is constant, so detailed balance is satisfied without additional corrections. On BPTI use of SCV led to about two orders of magnitude better performance than use of simple torsion angle MC. Recent work tested the use of SCV on a DNA oligomer [36] and concluded that its development "can lead to an interesting alternative to MD".

An other approach, mentioned already by Go and Scheraga [37], involves moving only the minimal set of torsion angles in the middle of a chain in such a way that the rest of the chain remains unmoved. This limits the atomic displacements resulting from the torsion angle changes and thus greatly enhances the efficiency of moves. For systems on a lattice the so-called pivot algorithm [38] implements such moves relatively easily. For off-lattice systems, however — besides the non-trivial geometry problem (solved first by Go and Scheraga [37]) — the formulation of a correct MC algorithm is quite challenging. The solution developed by Dodd et al. [39] has two salient points. First, the geometry problem of finding the torsion angles that keep the chain intact after the change of the first torsion has a variable number of solutions, thus the ratio q_{ij}/q_{ji} has to be proportional to the ratio of the number of solutions (when one of these solutions is chosen randomly for the trial move). Second, due to the constraints imposed on the set of torsion angles changed in each move, the volume element $d\phi_1 d\phi_1 \ldots d\phi_7$ is not constant, thus — as discussed above in Sec. 2 — the Boltzmann probability corresponding to each set of such torsion angles has to include the Jacobian of the change of coordinates from torsion angles to the constraint variables [39].

Hoffmann and Knapp [40] have suggested to improve the sampling efficiency by selecting a solution from the solution set with a probability that is proportional to the Jacobian. Additional improvement can be obtained, however, if one always selects the solution that is *nearest* to the previous state but, to maintain detailed balance, the change is rejected outright if the previous state is *not* the solution nearest to the new state [41].

3.5. Non-physical changes

Perhaps the strongest point of the MC methods is their ability to easily make discontinuous or other 'non-physical' changes. Thus, for systems with discontinuous variables (*e.g.,* lattice models) MC is generally the method of choice. Non-physical changes, however, can be very helpful in the sampling of the configuration space of continuous systems too. This section will present a collection of such successful techniques.

The method called configuration bias MC was developed from the work of Rosenbluth and Rosenbluth [42] on a method to generate self-avoiding random walks by Livne and Meirovitch [43] and, independently, by Siepman and Frenkel [44] for lattice simulations and by de Pablo et al. [45] for continuous models. In this approach, (part of) a chain molecule is rebuilt at each move segment by segment. To place a new segment, a decision about the orientation of the new bond (*i.e.,* the torsion angle around the last bond) is needed. This decision is based on the distribution of the Boltzmann factors as a function of this torsion angle, as estimated from the energies of a small sample of conformations. by biasing the selection of torsion angles toward values with high Boltzmann factor, the likelihood of building a new conformation of the polymer without overlapping with the rest of the system (or with itself) is greatly increased, even when the biasing factor is taken into account in the acceptance probability, Eq. (8). While the probability of successful completion of a chain growth decreases with the length of the chain the technique was successful even for lipid bilayers [46].

An other family of chain growing methods, called 'enrichment methods' grew out of the work of Wall and Erpenbeck [47]. These techniques involve multiple extension trys at promising steps of chain growing. The 'breadth first' strategy of Ref. 47 was replaced by a 'depth first' strategy in the works of Grassberger [48] and Hegger and Grassberger [49] and further enhancement by including a 'pruning' step (abandoning growing a chain when their weight falls below a threshold) by Grassberger [50]. Note, that these methods have a memory of the earlier steps of chain growing, thus the chain generated is non-Markovian.

A technique, similar in spirit to the chain growing with configurational bias, called slithering snake has been developed by Wall and Mandel [51] for lattice models and extended for continuum models by Kalos and coworkers (using the name 'reptation') [52,53]. The method simulates polymer melts including a so-called reptation step: one polymer unit is deleted from one end and an other is added (in random direction) at the other end. Note, that this method is limited to homopolymers.

Simulation in the grand-canonical ensemble (GCE) presented an other challenge, since it requires fluctuating density under constant volume, hence fluctuating number of atoms or molecules. The formalism generally used for GCE simulations has been developed by Adams [54]. Its application to dense fluids has been facilitated by the development of cavity-biasing where insertion attempt are made only at sites that are centers of a cavity of suitable radius [55,56]:

$$P_{\mathrm{ins}}^{\mathrm{acc}} = \min\{1, \exp(-\Delta E/kT + B)\frac{P_N^{\mathrm{cav}}}{N+1}\} \tag{11}$$

$$P_{\mathrm{del}}^{\mathrm{acc}} = \min\{1, \exp(-\Delta E/kT - B)\frac{N}{P_N^{\mathrm{cav}}}\}\ . \tag{12}$$

Here N is the number of particles, ΔE is the energy change upon the insertion or deletion, P_N^{cav} is the probability of finding a cavity in the current configuration, and B is the parameter introduced by Adams, related to the excess chemical potential μ' as

$$B = \mu'/kT + \ln\langle N\rangle\ . \tag{13}$$

Note, that P_N^{cav} is a byproduct of the cavity search, thus its incorporation into the q_{ij}'s incurs no additional computational cost.

GCE simulation of fluids is of particular interest as it yields the free energy without any extra work that otherwise can be quite substantial (see, *e.g.,* Ref. 57). For biomolecular systems it provides the unique capability of adequately filling internal cavities with water and eliminating waters improperly placed during the preparation of the system — see Ref. 58 for a costly example of such occurrence.

An analogue of the insertion/deletion employed in GCE simulations is the swapping of selected molecules, groups, or atoms. However, the author is not aware of any systematic exploration of this option.

A different generalization of the insertion/deletion process occurs in several variants of free-energy simulations [57]. Such calculations strive for finding the shortest path in the configuration space between the two states involved and therefore often lead to nonphysical path, usually best handled by MC.

Elimination of the requirement that attempted changes have to represent a physically meaningful process also greatly simplifies the treatment an arbitrary set of degrees of freedom differently from the rest. The simplest example for this is the freezing of an arbitrary set of degrees of freedom. Imposition of hard limits can be done also very simply. Other potentially useful 'non-physical' processes can include a simulated annealing where some degrees of freedom can be governed by a different temperature value.

A drawback common to MC and MD is the fact that the correlations between successive states die off too slowly. Simply abandoning a walk to start from a new random start is generally inefficient, however, since the equilibration time is usually longer than the correlation time. An intermediate solution to starting a different walk, called J-walking, has been introduced by Franz et al. [59]. In the simplest case, J-walking performs simulations at two temperatures simultaneously. Periodically, the simulation at the lower temperature attempts a 'move' that consist of switching from the lower temperature configuration to the one at higher temperature. The difference between the two temperatures are chosen to result in an overlap between the energy space of the two runs, resulting in 'reasonable' acceptance rates of such moves. However, the higher temperature run is expected to evolve much quicker than the lower temperature run, thus there is a reasonable probability that after an accepted switch the low temperature run is in a different basin of the configuration space. Further, J-walking does not have to be restricted to two temperatures. The higher temperature run can also be enhanced by a parallel run at an even higher temperature, and so on. The number of temperatures required is determined by the heights of the barriers between the conformational basins that the simulation is required to cross.

3.6. Generalized ensembles

In the last decade a family of simulation techniques emerged that perform simulations in so-called generalized or multicanonical ensembles. These methods have two conceptual precursors, simulated annealing and umbrella sampling. Simulated annealing was introduced as a method of finding the global minimum of a potential surface by Kirkpatrick et al. [60], performing Metropolis walks in the search space by progressively lower temperatures, while umbrella sampling was developed as a sampling enhancer for free energy simulations [13,14].

Note also that while they have been implemented in conjunction with the MC methodology, most of them could be implemented using MD.

The statistical mechanical foundation of the generalized ensembles follow from writing the configurational part of the (canonical) partition function as

$$Z = \int_{-\infty}^{+\infty} n(E) \exp(-E/kT) dE \tag{14}$$

where $n(E)\, dE$ is the volume of the configuration space with energy E. This implies that the probability that a random sample drawn from the Boltzmann distribution at temperature T Kelvin is proportional to $n(E) \exp(-E/kT)$. The exponential factor is 'responsible' for limiting the energies visited by a simulation at a given temperature. Simulated annealing simply drives the temperature from a high value toward zero, resulting in the sampling of a wide range. In the generalized ensembles, the simulation includes an additional weighing function to result in uniform sampling of a thermodynamic variable.

The method of expanded ensembles of Lyubartsev et al. [61] and the simulated tempering of Marinari and Parisi [62] includes the additional weight $\exp(-g(T))$ where uniform sampling in the *temperature* space results when $g(T) = F(T)/kT$, $F(T)$ being the Helmholtz free energy at temperature T. The multicanonical approach of Berg [63] includes the additional weight $w(E)$ where uniform sampling in the *energy* space results when $w(E) = 1/n(E)$. The additional weight $w(E)$ in the $1/k$ ensemble proposed by Hesselbo and Stinchombe [64] produces uniform sampling in the *entropy* space when $w_{1/k}(E) =\propto 1/\int_{-\infty}^{E} n(E')dE'$.

Hansmann and Okamoto gave a theoretical [65] and numerical comparison of these methods on the well-studied pentapetide met-enkephalin[66]. They found that these methods perform about equally well — a result understandable from the close relation between these techniques — and are superior to the traditional MC (or MD) techniques.

A common feature of these methods is that the ideal weight function is, while unknown, also the quantity targeted by the simulation (*i.e.,* the free energy), giving rise to an iterative scheme where trial calculations produce approximate forms to them, to be further improved by the successive simulations in an iterative fashion.

Such self-consistent algorithms have also their precursors in the umbrella sampling field (as noted in Ref. 64). While most free energy applications of umbrella sampling used a predetermined weighing scheme (*e.g.,* a harmonic function to concentrate sampling to a particular range of the free energy variable) adaptive methodology has also been used in umbrella sampling application for free energy simulations [67-69].

Andricioaei and Straub [70] showed the way to take advantage of the more delocalized character of the probability distribution in a different generalization of statistical mechanical ensembles [71]. This ensemble is derived from a generalized entropy, defined as a function of a parameter q, where the limit $q = 1$ gives the Boltzmann entropy. Note that this formalism can be also considered an application of umbrella sampling, but with the added advantage that the umbrella weight function is also defined in advance. They also combined the J-walking idea with this ensemble into a method called q-jumping: instead of trying a jump to a configuration reached at at different temperature, the jump attempt is made at a configuration generated with a different (higher) q value [72]. Further, Pak and Wang have implemented q-jumping with MD [73].

The generalized ensemble methods have been further extended by Wong and Liang [74] by introducing and importance weight associated with each state of the system (where the state may be in a generalized ensemble, like the ones discussed above). This weight is used to facilitate the crossing of barriers. The transition rules and the correct weighing of the states visited are also given in Ref. 74.

4. Optimization of move parameters

Besides the choice of the functional form of q_{ij} discussed above, the actual values of the parameters involved in q_{ij} can strongly effect the capabilities of a simulation to sample the configuration space. Foremost among the parameters is the so-called stepsize, the range within which the change in each coordinate should lie. Given that very small changes are almost always accepted but hardly move the system and very large steps are hardly ever accepted, early work suggested to tune the stepsize for 50% acceptance rate.

Optimization of stepsize parameters is a conceptually difficult proposition since it requires to estimate the efficiency of the walk generated by the simulation from the local properties of the walk. In other words, optimizations should be based on short runs to be of use. Typical such local properties are the mean square displacement and, for molecular liquids, orientational correlation. On this basis, Kincaid and Scheraga have demonstrated [75] that for liquid water acceptance rate of 30% results in significantly better sampling than the 50% rule. It should be pointed out that the sampling efficiency over long runs is more closely related to the limiting slope of the mean square displacement as a function of time [27] as well as to the 'time' scale of long-range correlations experienced in such systems [28] and it is an open question to see whether these latter quantities are proportional or even monotonic functions of the local convergence characteristics.

Depending on the implementation of the energy calculation, there may be an additional advantage to lower acceptance rates. Unless the program stores all pair interaction energies and related terms, the computational cost of a rejected move attempt is about half of the accepted moves. This follows from the need of updating the sum of the energies of all possible subsets that can be moved during the simulation only after accepted moves.

A detailed study on the optimal selection of the MC stepsize has been presented by Kolafa [76], again showing the virtues of acceptance rates significantly lower than 50%. Bouzida, Kumar and Swendsen [77] presented a systematic study of the relation between the acceptance rate and the stepsize and used it to devise a technique where the stepsize is periodically adjusted during the simulation. They note that this makes the whole chain a collection of different Markov chains (they call it almost Markov chain) but this fact does not appear to invalidate the method if the stepsize adjustment is not done very frequently.

5. Error estimates

The correlated nature of the configurations generated by the the Markov chain excludes the simple use of the standard deviation from being used for the determination of confidence intervals of the calculated averages. In a detailed review of the MC methodology, with plenty of attention to technical issues [78], Wood introduced to the simulation community the use of block averages. In this approach, the simulation is partitioned into blocks of sufficient length so that the averages over these blocks are already independent. In this case the standard deviation SD_X (corresponding to 68% confidence interval) of a property X is estimated as

$$SD_X = \sum_i^N \langle X \rangle_i / \sqrt{N-1} \tag{15}$$

where $\langle\,\rangle_i$ denotes the average of the quantity X over the i-th block and N is the number of blocks. The statistical independence of the $\langle X\rangle_i$'s has to be checked, however to ascertain the validity of Eq. (15). Also, the test can be repeated with increasing blocksizes and the convergence of the estimate can give an additional indication as to its reliability.

It is also possible to derive en estimate of the errors in the calculated averages by taking into account the correlated nature of the random walk. In this approach, the autocorrelation functions of the calculated quantities are also calculated. The autocorrelation approach can also be extended to the block-average approach — see, *e.g.,* Kolafa [79] for a detailed analysis of the problem and further references. The reliability of such estimates, however, hinges on the reliability of the estimated autocorrelations, creating a potential vicious circle.

6. Combining MC and MD

Given the complementary nature of the two techniques, it is natural to expect that combining the two can result in a method that is better than either alone. Such combination can be done either in the framework of MD or in that of MC.

Forrest and Suter [80] compared two versions of a MC method where each attempted change is generated by a short MD run, with initial velocities assigned from the Maxwell distribution. This combination was not found to perform better than MD for chains longer than 24 atoms.

The reverse approach, implementing MC steps into an MD calculation, appears to me more difficult, albeit likely to be of more use. The best results so far in this direction have been obtained by the method of Guarnieri [81] who introduced MC moves into a Brownian dynamics simulation. The success of the mixing hinges on the use of the so-called velocity-Verlet scheme to integrate the equations of motion since in that scheme dos not use the velocities of the previous steps. While formally not proven to be exact, the calculated distributions agreed well in every detail considered with results obtained with exact algorithms, but displayed significantly faster convergence. An other instance of MC steps mixed into and MD calculations is found in the work of Scott [82] who intersperses CBMC steps into an MD simulation. This solution is not an exact procedure either since the MC step is followed by a short minimization.

7. Implementation issues

While the implementation of the basic Metropolis method for atomic fluids is a rather simple matter, the more complex the system and the choice of the transition probabilities q_{ij}, the more involved the computer code becomes. Verifying the correctness of such algorithm and code is an important issue. Such verification is made difficult by the stochastic nature of the algorithm — there is no simple equation into which the result can be substituted to see if it is satisfied. However, novel algorithms and codes can be tested on systems where the answer is known. For example, Hofmann and Knapp tested their correlated torsion moves (called 'window moves' there) by simulating a system with no interaction thus the results of the simulation could be compared with the behavior established theoretically from the properties of uniform distribution.

The general lack of exact tests calls for consistency checks. Since the majority of MC methods involve changing small subsystems at a time, most calculated quantities are updated at each step instead of calculated from scratch. Already Wood suggested [78] the periodic comparison of the energy with the value recalculated from scratch. This idea has been further extended [83] and implemented with additional consistency checks on structural parameters in the program MMC [84]. This battery of consistency checks proved to be an invaluable help in detecting subtle errors in the code (some manifesting itself only after several millions of MC steps).

The idea of consistency checks can be generalized to compare properties calculated in different ways. Free energy simulation calculations are frequently checked by using twi different formalism or by repeating the calculation using a different path. A sensitive consistency test has been introduced by Butler et al. [85] who suggested the comparison of configurational termeprature with the temperature used in the acceptance probability. The calculation of configurational temperature requires the calculation of forces on the atoms, normally not required for MC simulation. Note, however, that if force biased displacements are used then using this test incurs no significant additional comutational burden.

The particularities of MC algorithms also give rise to novel opportunities for implementation using multiple processors. For example, in systems large enough to contain non-intersecting cutoff regions, moves of such far apart subsystems can be performed in parallel since the calculated ΔE will not be affected by the other change(s). Even for smaller systems where the independence of the ΔE's can not be ensured, the parallel calculation of several moves can still be an efficient coarse-grained paralellization, as follows.

1. Calculate ΔE_i for $i = 1, \ldots, n$ different attempted changes in parallel.

2. For each i, decide on acceptance in the usual way. If the move i is accepted, then for all j, $i < j \leq n$ update ΔE_j with the change in the energy between subsystems i and j. This step, however, can *not* be done in parallel.

Clearly, with small n, step 2 will be still fast, especially as low acceptance rates were found to be in general most efficient.

The implementation of fine-grained parallelism can also benefit from the fact that only a small part of the system is changed since that reduces the amount of data changed, lessening the communication load. Actual implementations, however, are strongly dependent on the actual hardware and software environment and their discussion is beyond the scope of this paper.

8. Concluding remarks

It is hoped that the various MC methodologies discussed briefly above gave a flavor for the many possibilities the MC methodology has. It is also clear that there are many possibilities for further enhancements, calling for additional development effort.

It is also important to emphasize that several techniques can be applied simultaneously for best effect. For example, reptation and concerted torsion angle moves were found to be a powerful combination [39]. It is also possible that the combination of enhanced sampling techniques results in synergy — this was found to be the case when force biasing [27] and preferential sampling [20] was combined [86].

Finally, it is important to stress that while the discussion above pointed out the aspects of MC that hold an advantage over MD, the final goal should be the development of such a modeling environment where the two techniques coexist, each is employed for tasks it is best suited. To reach such a state, however, the relative merits of the two approaches would have to be better defined — a difficult problem since the answer is system dependent — and methods have to be developed for the seamless integration of MD and MC.

Acknowledgement

Dr. Haluk Resat and the two anonymous referees are thanked for detailed constructive suggestions to the manuscript.

References

1. M.P. Allen and D.J. Tildesley, Computer simulation of liquids, Clarendon Press, Oxford (1986).

2. D. Frenkel and B. Smit, Understanding Molecular Simulations: from Algorithms to Applications, Academic Press, San Diego (1996).

3. N.A. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, Equation of state calculation by fast computing machines, *J. Chem. Phys.*, **21,** 1087–1092 (1953).

4. A. Rahman, Correlations in the motion of atoms in liquid argon, *Phys. Rev.,* **136,** A405–411 (1964).

5. F.H. Stillinger and C.W. David, Polarization model for water and its ionic dissociation products, *J. Chem. Phys.*, **69,** 1473–1484 (1978).

6. P. Jedlovszky and J. Richardi, Comparison of different water models from ambient to supercritcal conditions: A Monte Carlo simulation and molecular Orstein-Zernike study, *J. Chem. Phys.*, **110,** 8019–8031 (1999).

7. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.*, **98,** 10089–10092 (1993).

8. R. Sperb, An alternative to Ewald sums, part 2: The Coulomb potential in a periodic system, *Molecular Simulation,* **22,** 199–212 (1999).

9. W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57,** 97–109 (1970).

10. A.A. Barker, Monte Carlo calculations of the radial distribution functions for a proton-electron plasma, *Austral. J. Phys.*, **18,** 119–133 (1965).

11. P.H. Peskun, Optimum Monte Carlo sampling using Markov chains, *Biometrika*, **60,** 607–612 (1973).

12. J.S. Liu, F. Liang and W.H. Wong, The use of multiple-try method and local optimization in Metropolis sampling, *J. Amer. Statist. Assoc.,* **95,** 121–134 (2000).

13. G. Patey, and J.P. Valleau, Monte Carlo method for obtaining the interionic potential of mean force in ionic solution, *J. Chem. Phys.*, **82,** 2334 (1975).

14. G.M. Torrie, and J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comp. Phys.*, **23,** 187 (1977).

15. D. Levesque, J.J. Weis, and J.P. Hansen, Recent developments in the simulation of classical fluids, in "Applications of the Monte Carlo method in statistical physics", K. Binder, ed., Springer, Berlin, Heidelberg (1984).

16. Advances in Chemical Physics, **105,** "Monte Carlo methods in chemical physics" Wiley, New York, (1999).

17. MCMC Preprint server: http://www.statslab.cam.ac.uk/∼mcmc/

18. Molecular Monte Carlo home page: http://www.cooper.edu/engineering/chemechem/monte.html

19. M. Mezei, On the selection of the particle to be perturbed in the Monte Carlo method, *J. Comp. Phys.,* **39,** 128–136 (1981).

20. J.C. Owicki, Optimization of sampling algorithms in Monte Carlo calculations of fluids, in "Computer modeling of matter", P.G. Lykos, ed., American Chemical Society, Washington, D.C. (1978).

21. H. Resat, T.J. Marrone, and J.A. McCammon, Enzyme-inhibitor association thermodynamics: Explicit and continuum solvent studies, *Biophys. J.*, **74,** 2786–2801 (1998)

22. P. Jedlovszky, M. Mezei, Grand canonical ensemble Monte Carlo simulation of a lipid bilayer using extension biased rotations, *J. Chem. Phys.*, **111,** 10770–10773 (2000).

23. S. Goldman, A simple way to help speed up Monte Carlo convergence rates: Energy-scaled displacement Monte Carlo, *J. Chem. Phys.*, **79,** 3938–3947 (1983).

24. M. Mezei, K.A. Bencsáth, S. Goldman, and S. Singh, The detailed balance energy-scaled displacement Monte Carlo algorithm, *Molecular Simulation,* **1,** 87–93 (1987).

25. J. Cao and B.J. Berne, Monte Carlo methods for accelerating barrier crossing: Anti-force-bias and variable step algorithms, *J. Chem. Phys.,* **92,** 1980–1985 (1990).

26. H. Gordon, and S. Goldman, Simulations on the Primitive Electrolyte Environment of a High Charge-Density Polyelectrolyte. A Sampling Problem and its Solution, *Molecular Simulation,* **3,** 213–225 (1989).

27. R.M. Rao, C.S. Pangali, and B.J. Berne, On the force bias Monte Carlo simulation of water: Methodology, optimization and comparison with molecular dynamics, *Molec. Phys.,* **37,** 1779–1798 (1979).

28. M. Mezei, S. Swaminathan, D.L. Beveridge, Convergence characteristics of the conventional Metropolis Monte Carlo method on liquid water, *J. Chem. Phys.,* **71,** 3366–3373 (1979).

29. M. Mezei, Virial-bias Monte Carlo methods: Efficient sampling in the (T, P, N) ensemble, *Molec. Phys.,* **48,** 1075–1082 (1983).

30. M. Mezei, Distance-scaled force biased Monte Carlo simulation for solutions containing a strongly interacting solute, *Molecular Simulation,* **5,** 405–408 (1991).

31. B.J. Hardy, and R.W. Pastor, Conformational sampling of hydrocarbon and lipid chains in an orienting potential, *J. Comp. Chem.*, **15,** 208–226 (1994).

32. F. Guarnieri and H. Weinstein, Conformational memories and the exploration of biologically relevant peptide conformations: An illustration for the gonadotropin-releasing hormones, *J. Am. Chem. Soc.*, **118,** 5580–5589 (1996).

33. G. Némethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H.A. Scheraga, Energy parameter in polypeptides. 10. Improved geometrical parameters

and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides, *J. Phys. Chem.*, **96,** 6472–6484 (1992).

34. W.L. Jorgensen, and J. Tirado-Rives, Monte Carlo vs molecular dynamics for conformational sampling, *J. Phys. Chem.*, **100,** 14508–14513 (1996).

35. T. Noguti and N. Go, Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins, *Biopolymers*, **24,** 527–546 (1985).

36. H.A. Gabb, C. Prevost, G. Bertucat, C.H. Robert, R. Lavery, Collective-variable Monte Carlo simulation of DNA, *J. Comp. Chem.*, **18,** 2001–2011 (1997).

37. N. Go, and H.A. Scheraga, Ring closure and local conformational deformation of chain molecules, *Macromolecules*, **3,** 178–187 (1970).

38. A. Kolinski, J. Skolnick, Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme, *Proteins,* **18,** 338-352 (1994).

39. L.R. Dodd, T.D. Boone, and D.N. Theodoru, A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses, *Molec. Phys.,* **78,** 961–996 (1993).

40. D. Hoffman, and E.-W. Knapp, Polypeptide folding with off-lattice Monte Carlo dynamics: The method, *Eur. Biophys. J.*, **24,** 387–403 (1996).

41. M. Mezei, submitted.

42. M.N. Rosenbluth, and A.V. Rosenbluth, Monte Carlo calculation of the average extension of a molecular chain, *J. Chem. Phys.*, **23,** 356–359 (1955).

43. S. Livne, and H. Meirovitch, Computer simulation of long polymers adsorbed on a surface. I. Correction to scaling in an ideal chain. *J. Chem. Phys.*, **88,** 4498–4506 (1988).

44. J.I. Siepmann, and D. Frenkel, Configurational bias Monte Carlo: A new sampling scheme for flexible chains, *Molec. Phys.*, **75,** 59–76 (1992).

45. J.J. de Pablo, M. Laso, and U.W. Suter, Simulation of polyethylene above and below the melting point, *J. Chem. Phys.*, **96,** 2395–2403 (1992).

46. H.L. Scott, E. Jacobsson, and S. Subramaniam, Simulations of lipid membranes with atomic resolution, *Comput. Phys.* **12,** 328–334 (1998).

47. F.T. Wall and J.J. Erpenbeck, New method for the statistical computation of polymer dimensions. Statistical computation of radii of gyration and mean internal dimensions of polymer molecules. *J. Chem. Phys.,* **30,** 634–637 (1959).

48. P. Grassberger, Recursive sampling of random walks: Self-avoiding walks in disordered media, *J. Phys. A,* **26,** 1023–1036 (1993).

49. R. Hegger and P. Grassberger, Chain polymers near adsorbing surface, *J. Phys. A,* **27,** 4069–4081 (1994).

50. P. Grassberger, Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000, *Phys. Rev. E,* **56,** 3682–3693 (1997).

51. F.T. Wall and F. Mandel, Macromolecular dimensions obtained by an efficient Monte carlo method without sample attrition, *J. Chem. Phys.*, **63,** 4592–4595 (1975).

52. I. Webman, J.L. Lebowitz and M.H. Kalos, MOnte Carlo studies of a polymer between plates, crossover between dimensionality, *J. Phys. (Paris),* **41,** 579–583 (1980).

53. M. Bishop, D. Ceperly, H.L. Frisch, and M.H. Kalos, Investigations of static properties of model bulk polymer fluids, *J. Chem. Phys.*, **72,** 3228–3235 (1980).

54. D.J. Adams, Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid, *Molec. Phys.*, **29,** 307–311 (1975).

55. M. Mezei, A cavity-biased (T,V,μ) Monte Carlo method for the computer simulation of fluids, *Molec. Phys.,* **40,** 901–906 (1980).

56. M. Mezei, Grand-canonical ensemble Monte Carlo simulation of dense fluids: Lennard-Jones, soft spheres and water, *Molec. Phys.,* **61,** 565–582 (1987); Erratum, **67,** 1207–1208 (1989).

57. M. Mezei, and D.L. Beveridge, Free energy simulations, *Ann. Acad. Sci N.Y.,* **482,** 1–23 (1986).

58. D.P. Tieleman, and H.J.C. Berendsen, A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyloleylphospatidycholine bilayer. *Biophys. J.*, **74,** 2786–2801 (1998).

60. S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi, Optimization by simulated annealing *Science,* **220,** 671–680 (1983).

61. Q.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov and P.N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles, *J. Chem. Phys.,* **96,** 1776–1783 (1992).

62. E. Marinari and G. Parisi, Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters,* **19,** 451–455, 1992

63. B.A. Berg and T. Neuhaus, Multicanonical ensemble — a new approach to simulate 1st order phase transitions, *Phys. Rev. Lett.,* **68,** 9–12 (1992).

64. B. Hesselbo and R.B. Stinchcombe, Monte Carlo simulation and global optimization without parameters *Phys. Rev. Lett.* **74,** 2515–2155 (1995).

59. D.D. Franz, D.L. Freeman and J.D. Doll, reducing quasi-ergodic behavior in Monte Carlo simulations by J-walking: Application to atomic clusters, *J. Chem. Phys.,* **93,** 2769–2784 (1979).

65. U.H.E. Hansmann and Y. Okamoto, Monte Carlo simulations in generalized ensemble: Multicanonical algorithm versus simulated tempering, *Phys. Rev. E,* **54,** 5863–5865 (1996)

66. U.H.E. Hansmann and Y. Okamoto, Numerical comparisons of three recently proposed algorithms in the protein folding problem, *J. Comp. Chem.*, **18,** 920–933 (1997).

67. G.M. Paine, and H.A. Scheraga, Prediction of the native conformation of a polypeptide by a statistical-mechanical procedure. I. Backbone structure of enkephalin, *Biopolymers*, **24,** 1391–1436 (1985).

68. M. Mezei, Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias, *J. Comp. Phys.*, **68,** 237–248 (1987).

69. C. Bartels and M. Karplus, Multidimensional Adaptive Umbrella Sampling: Applications to main chain and side chain peptide conformations, *J. Comp. Chem.,* **18,** 1450–1462 (1997).

70. I. Andricioaei anf J.E. Straub, Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide, *Phys. Rev. E,* **53,** R3055–R3058 (1996).

71. D.A. Stariolo and C. Tsallis, *Annual Reviews of Computational Physics II,* D. Staufer, ed., World Scientific, Singapore (1995) p 343.

72. I. Andricioaei anf J.E. Straub, On Monte Carlo and molecular dynamics methods inspired by Tsallis statisistics: Methodology, optimization, and application to atomic clusters, *J. Chem. Phys.,* **107,** 9117–9124 (1997).

73. Y. Pak and S. Wang, Folding of a 16-residue helical peptide using molecular dynamics simulation with Tsallis effective potential, *J. Chem. Phys.,* **111,** 4359–4361 (1999).

74. W.H. Wong and F. Liang, Dynamic weighing Monte Carlo and optimization, *Proc. Natl. Acad. Sci. USA,* **94,** 14220–14224 (1997).

75. R.H. Kincaid, and H.A. Scheraga, Acceleration of convergence in Monte Carlo simulations of aqueous solutions using the Metropolis algorithm. Hydrophobic hydration of methane, *J. Comp. Chem.*, **3,** 525–547 (1982).

76. J. Kolafa, On optimization of Monte Carlo simulations, *Molec. Phys.*, **63,** 559–5792 (1988).

77. D. Bouzida, S. Kumar, and R.H. Swendsen, Efficient Monte Carlo methods for the computer simulation of biological molecules, *Phys. Rev. A.*, **45,** 8894–8901 (1992).

78. W.W. Wood, Monte Carlo studies of simple liquid models, in "Physics of simple liquids." H.N.V Temperly, F.S. Rowlinson, and G.S. Rushbrooke, Eds. North-Holland (1968).

79. J. Kolafa, Autocorrelations and subseries averages in Monte Carlo simulations, *Molec. Phys.*, **59,** 1035–1042 (1986).

80. B.M. Forrest, and U.W. Suter, Hybrid Monte Carlo simulations of dense polymer systems *J. Chem. Phys.,* **101,** 2616–2629 (1994).

81. F. Guarnieri, Theory and algorithms for mixed Monte Carlo – stochastic dynamics simulations, *J. Math. Chem.,* **18,** 25–35 (1995).

82. S.-W. Chiu, M. Clark, E. Jakobsson, S. Subramaniam and H.L. Scott, Application of a Combined Monte Carlo and Molecular Dynamics Method to the Simulation of a Dipalmitoyl Phosphatidylcholine Lipid Bilayer, *J. Comp. Chem.,* **20,** 1153–1164, (1999).

83. M. Mezei, A Comment on debugging Monte Carlo programs, *Information Quarterly, CCP5,* Daresbury Lab., No **23,** 33–37 (1986).

84. M. Mezei, MMC: Monte Carlo program for simulation of molecular assemblies. URL: http://inka.mssm.edu/∼mezei/mmc

85. B.D. Butler, G. Ayton, O.G. Jepps, and D.J. Evans, Configurational temperature: Verification of MOnte Carlo simulatios, *J. Chem. Phys.,* **109,** 6519–6522 (1998).

86. P.K. Mehrotra, M. Mezei, and D.L. Beveridge, Convergence Acceleration in Monte Carlo Computer Simulation on Water and Aqueous Solutions, *J. Chem. Phys.,* **78,** 3156–3166 (1983).