

Efficient molecular simulations for environmentally benign processes

S. ULAS^{†‡} and U. M. DIWEKAR^{†‡*}

[†]Department of Bioengineering, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607, USA

[‡]Vishwamitra Research Institute, Center for Uncertain Systems: Tools for Optimization and Management, 34 N. Cass Avenue, Westmont, IL 60559, USA

(Received December 2005; in final form January 2006)

An efficient sampling technique namely the Hammersley sequence sampling (HSS) technique based on quasi-random numbers is applied to molecular simulations to increase computational efficiency and reduce the error in property calculations for efficient and environmentally friendly product and process design. Due to its k -dimensional uniformity properties, this sampling technique reduces the computational intensity and converges faster as compared to conventional Monte Carlo (MC) method based on pseudo-random numbers. Two case studies are presented in order to show the effect of the sampling technique on molecular simulations. The first case study involves the computation of equation of state (EOS) for a Lennard Jones fluid and the second case study concerns the prediction of Gibbs free energies of transfer for the calculation of octanol–water partition coefficients. The results show that the HSS technique provides about 3-fold gain in efficiency and reduces the number of cycles required to reach equilibration. Furthermore, the errors in property calculations are reduced for both of the case studies.

Keywords: Octanol–water partition coefficients; Equations of state; Hammersley sequence sampling (HSS); Quasi-Monte Carlo method

1. Introduction

Chemical process industries generate a lot of waste in the form of materials and energy, where valuable resources are lost. Furthermore, toxic waste and emissions pose a threat to human health and the environment. Substantial waste minimization and/or prevention can only be realized if environmental considerations are included in every aspect of process design. This includes an integrated approach for environmentally benign chemical selection/synthesis, process synthesis and design [1]. In order to successfully execute these design steps, reliable data is needed for important properties of chemicals.

The mathematical models needed for environmentally benign product and process design often require quantification of the environmental impact of chemicals, and other thermodynamic properties, which affect process performance indicators like efficiency, and flexibility. For these properties, the accuracy of prediction methods, as well as the availability of reliable data can be a problem. Molecular simulations can play a significant role in enhancing the accuracy (reducing uncertainties)

in prediction. Furthermore, when new chemicals are produced, where no experimental data for properties are available, they can provide predictive capabilities. Molecular simulations, not only allows us to predict the properties of a particular substance but also provides us with microscopic insight on the energetics and various mechanisms of the systems studied. The success of Monte Carlo (MC) and molecular dynamics simulations depends on the efficient sampling of the configuration space, to generate states of low energy, enabling the calculation of properties accurately. In MC simulations, particles are randomly selected and moved by a random extent and the energy change of the system is analyzed. For systems with large number of molecules, this task requires significant computational time. In order to reduce the computational intensity of molecular simulations, researchers have used many sampling techniques which introduce a bias to obtain samples from a region of importance. However, most of these techniques are system dependent and are not universally applicable. In this work, we are introducing a universal approach for increasing the efficiency of molecular simulations. In order to show the wider

*Corresponding author. Email: urmila@vri-custom.org

applicability of this approach, we have considered two types of property prediction models by molecular simulations, namely, thermodynamic property like phase equilibrium crucial for defining separation efficiency, and environmental impact property prediction.

1.1 Uncertainties in group contribution models for phase equilibria

Separation processes are crucial for removing toxic and hazardous substances from waste streams and play an important role in greener process design. Separation efficiency depends on phase equilibrium properties such as vapor–liquid equilibrium (VLE), which is crucial in solvent intensive processes. Although there is considerable amount of experimental data available for VLE [3], due to the amount of new compounds being produced in laboratories everyday, the experimental data available is not adequate and the models available especially for the early stages of design like chemical synthesis (or selection) have a lot of uncertainties.

For example, figure 1 shows the uncertainties in more than 1800 interaction parameters present in the UNIFAC model collapsed in terms of three parameters (organic–water, organic–organic and water–organic) to consider the difference between water and organic chemicals, in a study by Kim and Diwekar [4,5]. In this figure, uncertainty factor is defined as the ratio of experimental values of γ^∞ , the activity coefficient at infinite dilution to the γ^∞ values predicted by the UNIFAC model. The deviation from the value of 1 shows the uncertainties in this model based on group contributions. As can be seen from the figure, the value of uncertainty factor varies between 1 and 20, which affects the reliability of group contribution models in process design.

1.2 Models for predicting environmental impact

Toxicological properties such as lethal concentration LC_{50} and lethal dose LD_{50} , which would kill 50% of a test

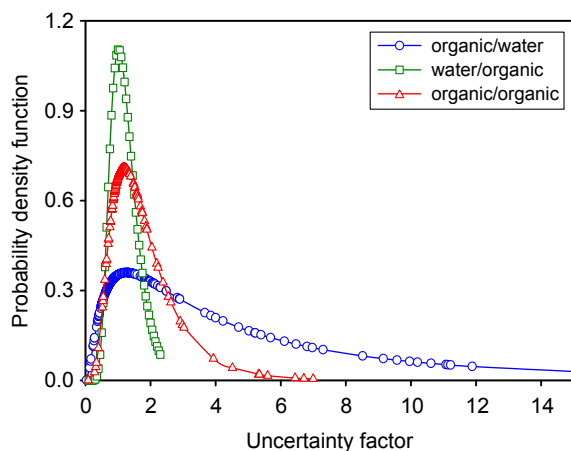


Figure 1. Uncertainties in UNIFAC model [4].

population of experimental animals, as well as bio-concentration factor, which describes the bio-accumulation of chemicals in an organism, are the basis for defining the environmental impacts of chemicals in a process. These toxicological parameters are closely related to the octanol–water partition coefficient, K_{ow} . It has been shown that bio-concentration factor has a strong dependence on K_{ow} and many direct relationships have been reported between LC_{50} , LD_{50} and K_{ow} [6]. Furthermore, octanol–water partition coefficients provide a basis to a large amount of quantitative structure activity relationships (QSARs). QSARs are used to correlate or predict solute properties in bio-phases (membranes, fat tissue, body fluids) to assess the biological activity of chemicals inside the body.

The octanol–water partition coefficient, K_{ow} in thermodynamics is a free energy function which is directly related to the energetics of transfer between two phases. The successfulness of K_{ow} in predicting bio-accumulation and toxicity have been attributed to the fact that the octanol–water partitioning is a good representative of a compound going from more aqueous like-phases (extracellular phase) to organic-like phases (cellular phase) inside the body, which is the rate controlling step during the interaction of a compound and a biological system [7]. Therefore, this coefficient is used for environmental risk assessment in a variety of different settings.

Similar to phase equilibrium data, experimental data is also available for octanol–water partition coefficient (<http://logkow.cisti.nrc.ca/logkow/coeff.html>). However, due to the increasing amount of new chemicals synthetically produced everyday, there is a growing demand for efficient methods to predict or correlate the toxicological properties.

Octanol–water partition coefficients are calculated using a variety of different methods such as fragment of group contribution methods, atomistic methods, whole molecule approaches, and linear solvation energy relationships. Similar to phase equilibria, octanol–water partition coefficients are also predicted using group contribution methods [8–11]. However, uncertainties are inherent in these models used to predict octanol–water partition coefficients.

According to the study of Lin and Sandler [10], the error in group contribution models ranges from 38 to 71%. This large error will affect the environmental risk assessment and the toxicological and biological properties predicted using K_{ow} . Also for most of these group contribution models, the groups needed to predict the properties for certain chemicals are not available, which also affects the accuracy of these models.

Furthermore, a recent study by Linkov *et al.* [12] points out the fact that there is considerable variability among the tabulated values for octanol–water partition coefficients and this means there are uncertainties in experimental data as well. In some cases, this is due to the small amounts of emulsified octanol in water phase, which results in erroneously calculated values of K_{ow} .

1.3 Efficient sampling approach to molecular simulations

MC methods for various property predictions use random numbers for particle selection and movements, which are based on pseudo-random numbers. In this work, in order to reduce the computational time and improve the efficiency of molecular simulations, we have used Hammersley sequence sampling (HSS) technique [13]. This sampling technique provides a generalized approach for increasing simulation efficiency by systematically replacing the pseudo-random numbers used in MC simulations with quasi-random numbers. This sampling technique uses low-discrepancy sequences (quasi-random points) which have the better uniformity or evenness in their domain of definition. This sampling technique has been shown to require fewer samples and faster convergence properties in various applications such as off-line quality control of a CSTR [13], robust design of distillation columns [14], solvent selection [4,5], multi-objective optimal designs for emission reduction [15] and optimal molecular design under uncertainty [16].

In this work, quasi-random points showing k -dimensional uniformity generated by the HSS technique are used instead of pseudo-random points for various MC moves of the molecular simulation algorithms for the prediction of octanol–water partition coefficients and phase equilibria (equation of state (EOS)), to obtain faster convergence and reduce the computational requirements. Using this technique, we are able to significantly reduce the number of MC cycles for accurate estimation of properties and improved the efficiency of equilibration.

This paper is divided into six parts. Sections 2, describes the importance of octanol–water partition coefficient and details of molecular simulations to predict octanol–water partition coefficients and equations of state. In Section 3, a literature survey is presented on various methods to increase the efficiency of molecular simulations and the current method proposed based on quasi-random numbers and the HSS technique is explained in detail. In Section 4, the application of this efficient sampling technique to molecular simulations for the prediction of K_{ow} and equations of state predictions is detailed. Then we compare the results we have obtained from pseudo-random points and HSS samples and provide a discussion. In Section 5, summary and conclusions are presented.

2. Molecular simulations for property predictions

Molecular simulations are used for the prediction of important thermodynamic and structural properties of chemicals such as EOS, transition temperature, coefficient of thermal expansion, heat capacities (C_p , C_v), free energy changes, molar volume and density. In this work, our focus is on EOS predictions and the estimation of octanol–water partition coefficients from Gibbs free energies of transfer using molecular simulations.

2.1 Equation of state predictions by molecular simulations

One of the most important applications of molecular simulations is to compute the phase diagram of a given model system. This can be achieved by investigating the EOS for a given system. Equations of state are used frequently to calculate the thermodynamic properties of real fluids. An EOS is a relation between the temperature, pressure, number of atoms and volume of the system studied.

The Lennard Jones fluid is studied most frequently and it provides a good description of liquid/vapor coexistence properties of spherical (argon, krypton, xenon) and homonuclear diatomic molecules. Lennard Jones model is also widely employed for studying the vapor/liquid coexistence in alkanes and polar molecules (water, methanol) where the atom–atom interactions are described by the Lennard Jones potential.

The full Lennard Jones potential can be approximated as follows:

$$u(r) = \begin{cases} u^{lj}(r), & r \leq r_c \\ 0, & r > r_c \end{cases} \quad (1)$$

where the cutoff radius r_c is set to half the box length. The contribution of the particles beyond this cutoff is estimated with the usual tail corrections.

For energy:

$$u^{\text{tail}} = \frac{8}{3} \pi \rho \left[\frac{1}{3} \left(\frac{1}{r_c} \right)^9 - \left(\frac{1}{r_c} \right)^3 \right] \quad (2)$$

For pressure:

$$p^{\text{tail}} = \frac{16}{3} \pi \rho^2 \left[\frac{2}{3} \left(\frac{1}{r_c} \right)^9 - \left(\frac{1}{r_c} \right)^3 \right] \quad (3)$$

The EOS for the Lennard Jones fluid has been studied by many groups using molecular dynamics or MC simulations. In a study by Frenkel and Smit [17] the Metropolis algorithm was used with basic particle moves. The pressure was calculated using the virial, vir , which is defined by:

$$\text{vir} = \frac{1}{3} \sum_i \sum_{j>i} f(r_{ij}) \cdot r_{ij} \quad (4)$$

$$P = \frac{\rho}{\beta} + \frac{\text{vir}}{V} \quad (5)$$

where $f(r_{ij})$ is the intermolecular force. Equation (5) then provided the EOS. In this work, we are using the approach proposed by Frenkel and Smit [17] for prediction of EOS.

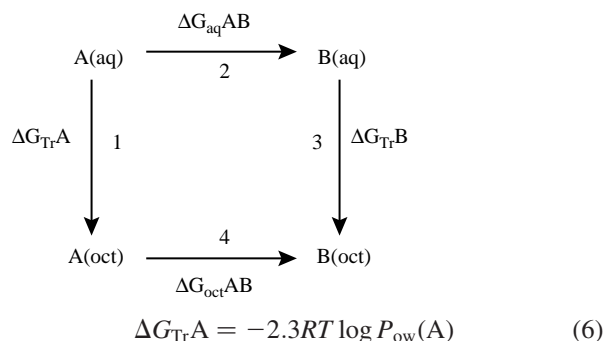
2.2 Octanol–water partition coefficient and prediction methods by molecular simulations

The second part of our case studies deals with the prediction of octanol–water partition coefficients. Octanol–water partition coefficient (K_{ow}), in thermodynamics is a free energy function (like solubility and vapor pressure) and, therefore, it is directly

related to the energetics of transfer between two phases. The connection of K_{ow} with biological activity was suggested by Hansch and Fujita [7], as shown in figure 2. This figure shows the interaction of a compound with a biological system. At first, the chemical compound arrives at a particular site in a cell from a dilute solution outside the cell following a random walk. Then the chemical compound goes through the relatively slow process of diffusion or permeation, which represents the partitioning of the compound between a polar aqueous phase and a non-polar organic phase. This partitioning is controlled by the molecular structure of the compound. Once the compound goes into the non-polar organic phase it goes through a series of chemical reactions eliciting a biological response. However, the rate-controlling step in this scheme is the first step, which is the partitioning of the compound between a lipophilic and hydrophilic phase and octanol–water partition coefficient is a good indicator of this step. Octanol has a structure, which is composed of a hydrophilic head and a lipophilic tail, which has been found to mimic the complexities of biological and other environments very well. This is the reason for its importance in estimating various toxicological parameters such as bio-concentration factor and LC_{50} , LD_{50} and its connection to QSARs as mentioned in previous section.

There are three different methods available in literature to calculate K_{ow} by molecular simulations. The first method is the free energy perturbation (FEP) method. Calculations of relative octanol–water partition coefficients have been reported using FEP method with the molecular dynamics and MC simulation methods. The computational strategy is as follows (Jorgensen *et al.* 1990; [58]):

If one considers the thermodynamic cycle below for two solutes A and B:



$$\Delta G_{Tr,B} = -2.3RT \log P_{ow}(B) \quad (7)$$

In equations (6) and (7), the logarithm of the octanol–water partition coefficients for solutes A and B are presented in terms of free energies of transfer. The difference between the solvation free energies is the free energy of transfer

$\Delta G_{Tr,A}$ for transferring solute A from water to water-saturated octanol. From the thermodynamic cycle, we can obtain equation (8), since we know the fact that free energy is a state function.

$$\Delta G_{Tr,B} - \Delta G_{Tr,A} = \Delta G_{oct}^{AB} - \Delta G_{aq}^{AB} = \Delta \Delta G_{Tr,AB} \quad (8)$$

$$\Delta \Delta G_{Tr,AB} = -2.3RT \Delta \log P_{ow} \quad (9)$$

Since it is difficult to calculate absolute free energies of solvation, a relative free energy of solvation is computed in which solute A is slowly mutated to solute B. Equation (9) allows the calculation of the relative partition coefficient ($\Delta \log P_{ow}$) for solutes A and B to be calculated from the direct relationship between the relative free energy of transfer $\Delta \Delta G_{Tr,AB}$ and $\Delta \log P_{ow}$.

Using the above mentioned methodology and FEP methods, molecular dynamics simulations were performed to calculate the relative octanol–water partition coefficients [2,18].

The second method to calculate K_{ow} is the expanded ensemble method. Recently, Lyubartsev *et al.* [19] have used the expanded ensemble method, which provides a realistic way of modeling the two phase system at specific physical conditions, within the framework of molecular dynamics simulations to compute octanol–water partition coefficients. The expanded ensemble method can both be used with a MC (molecular simulations) and molecular dynamics framework. Using this method, not only the relative partition coefficients but also absolute values for the partition coefficients can be obtained. This method involves the gradual insertion/deletion of studied solute particle into/from the solvent. One drawback of this method is that it requires optimization of two parameters, insertion parameters and balancing factors for the simulation, so some trial runs are needed.

The third method that is used to calculate octanol–water partition coefficients is the configurational-bias MC (CBMC) simulations in the Gibbs ensemble [20]. The Gibbs ensemble is very suitable for this simulation because it creates a setup analogous to the experimental situation. Gibbs ensemble Monte Carlo simulation (GEMC) utilizes two separate simulation boxes, which are in thermodynamic contact, but do not have an explicit interface. The advantage of this latter work to FEP studies is that the Gibbs free energy of transfer can be determined directly from the ratio of solute number densities whereas the difference in excess chemical potentials is used in FEP and the number density ratio can be determined very precisely from CBMC/GEMC simulations. Also the compositions of the two solvent phases do not need to be specified in advance.

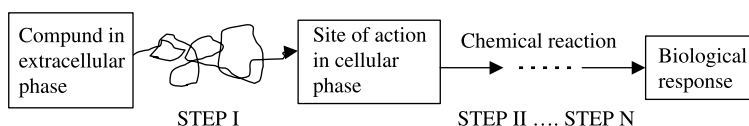


Figure 2. Simplified model for the interaction of a chemical compound and a biological system [57].

In this paper, we use this latter methodology for predicting octanol/water partition coefficients (K_{ow}) from Gibbs free energies of transfer. Our aim is to enhance the efficiency of predicting K_{ow} from molecular simulations by introducing an efficient sampling technique. Section 3 presents a review of sampling techniques to improve the efficiency of MC simulations and our methodology to tackle this problem.

3. Role of sampling in molecular simulations

The Metropolis Monte Carlo (MMC) method is the first approach in increasing efficiency of MC molecular simulations by using the “importance sampling” concept. In importance sampling, a biased distribution is used to obtain more samples from a region of importance. Boltzmann distribution function is such a distribution used in MMC where system configuration states which make substantial contributions to the ensemble averages are generated.

The MMC method requires large number of samples to generate accurate property estimations and is computationally intensive especially for large number of molecules and complex fluids. Therefore, many researchers have worked on sampling techniques in order to speed up the calculations and cover the configurational space more efficiently. Some of these sampling techniques are reviewed in the following subsection.

3.1 A survey of sampling techniques for efficiency improvement in Monte Carlo molecular simulations

The most commonly employed sampling techniques to overcome the limitations of conventional MC algorithm are biased sampling techniques. CBMC was introduced by Siepmann and Frenkel [21] to simulate flexible chain molecules. This scheme is based on a self-avoiding random walk where trial conformations are generated using a Rosenbluth scheme. Several extensions of this algorithm have been developed to improve its efficiency, such as dual-cutoff CBMC, where a shorter range cutoff is used [22], coupled–decoupled CBMC [23] where intramolecular terms were now also generated using a biasing procedure, self-adapting fixed-end-point CBMC [24] for efficient conformational sampling of the interior segments of chain molecules, rebridging CBMC [25] to simulate inner sections of linear and cyclic molecules and a new CBMC method for lattice polymers [26]. The final method that should be mentioned in biased sampling techniques is the aggregation-volume-bias (AVBMC) method [27], to enhance the efficiency of sampling the phase space of strongly associating fluids by introducing a novel MC move targeted at sampling the formation (growth) or destruction (shrinkage) of aggregates. All of these methods allow for MC trial moves with a much higher acceptance rate and, therefore, increase the efficiency of MC simulations, specifically grand canonical MC and Gibbs ensemble simulations.

Another class of sampling techniques to improve the efficiency of MC and molecular dynamics simulations use non-Boltzmann biasing. These algorithms are called generalized ensemble algorithms, where each state is weighted by an artificial, non-Boltzmann probability weight factor so that a random walk in potential energy space may be realized. The random walk allows the simulation to escape from any energy barrier and to sample much wider configurational space than by conventional methods. A review of these methods can be found in reference [28,29].

The most well known method of this kind is the multicanonical algorithm (MUCA) [30]. This method is also called entropic sampling or adaptive umbrella sampling. In this method, the potential function is modified using a non-Boltzmann weighting function so that the unfavorable states are sampled sufficiently. A free 1D random walk in potential energy space is realized with this method. This allows the simulation to escape from any local minimum energy states and sample the configuration space much more widely. The second method that belongs to generalized ensemble algorithms is the simulated tempering (ST) [31,32] method which is also called method of expanded ensemble. Differing from MUCA this method performs a free 1D random walk in temperature space. This induces a random walk in potential energy space which allows the simulation to escape from local energy states. The third method that belongs to the group of generalized ensemble algorithms is the replica-exchange method (REM) [33], which is also referred to as multiple Markov chain method and parallel tempering. In ST and umbrella sampling, probability weight factors need to be known beforehand and they are often determined by short trial simulations. However, with REM this process is not needed because with this method, non-interacting replicas of the original system at different temperatures are simultaneously and independently simulated with the conventional MC or MD method and every few steps these replicas are exchanged with a specified transition probability. The weight factor is just the product of Boltzmann factors, and so it is essentially known. However, this method is computationally very expensive as the number of degrees of freedom increases. Several publications exist in the literature about these algorithms and their variations and they are often used to simulate complex systems such as spin glass and biopolymers.

One other method to improve the sampling efficiency of MC simulations of systems containing large potential energy barriers is J-walking. This method makes possible the jumping over these barriers by coupling the usual Metropolis sampling to the Boltzmann distribution generated by another random walker at a higher temperature. This method shortens the time required to sample conformation space. This method was initially applied to atomic clusters [34]. Zhou and Berne [35] developed a variant of this method called Smart Walking (S-walking), which uses relaxed structures to study

proteins. A recently developed stochastic potential switching algorithm [36] uses a similar principle by replacing the actual potential V with a less complex potential to reduce the complexity of interactions among the particles.

Several other methods of increasing the sampling efficiency of MC simulations exist in literature. Some examples are, valley restrained MC method which predicts the topology of the conformational energy hyper-surface using statistical and empirical data for proteins, as a method to improve the sampling efficiency [37], preferential cell sampling method [38] used in grand canonical ensemble for dense and confined fluids, which is reported to be several orders of magnitude more efficient than crude MC sampling, and 4–40 times more efficient than restricted umbrella sampling, the adiabatic nuclear and electronic sampling Monte Carlo algorithm (ANES-MC) [39], which provides additional sampling of the electronic degrees of freedom to improve the efficiency of particle swap moves, parallelized sampling of Gibbs ensemble [40] which is based on an expression for the probability distribution of NVT-like sub-ensembles, extended state space MC method [41], which attempts to increase the sampling mobility by introducing new parameters and a statistical quenching and heating procedure similar to simulated annealing and a method which is based on proper sampling of rejected states to speed-up MC simulations [42].

Information theory and genetic algorithms have also been used in order to optimize the sampling efficiency of MC simulations. Recently, Leblanc *et al.* [43] presented a new approach in order to improve the convergence of MC simulations of molecular systems belonging to complex energetic landscapes. This method dynamically allocates MC move frequencies depending on their past efficiency, measured with respect to a relevant sampling criterion and this is achieved using an evolutionary algorithm, turning the sampling efficiency of molecular simulation into an optimization problem. Significant improvement in sampling efficiencies was observed using this method. Alternatively, Eleftheriou *et al.* [44] developed a quantitative measure of the sampling efficiency of equilibrium MC algorithms and based on information theory this method was optimized by independent random sampling. Despite the fact that, the literature is abundant in examples of algorithms for efficiency improvement for MC simulations, most of these methods are derived based on the importance sampling (biased sampling) principle and are often problem specific and offer customized solutions for specific systems.

In this study, we propose a universal approach of improving efficiency. All MC simulation methods use random numbers for their analysis. The approach here is to replace some of these random numbers with quasi-random numbers generated by an efficient sampling technique called HSS. Then k -dimensional uniformity property of this sampling is exploited when replacing these pseudo-random numbers. It has been found in an earlier work related to simulated annealing [4,5] that it is important to differentiate where the k -dimensional uniformity is

important. We compare the results obtained by pseudo-random numbers and quasi-random numbers for the estimation of two properties in order to show that this method is universally applicable for any property model predictions. It should be noted that this proposed approach can be used in conjunction with the biased MCS strategies presented in literature and is not restricted to specific applications.

3.2 Hammersley sequence sampling

As mentioned above, conventional MC methods use pseudo-random numbers in their calculations to approximate a uniform distribution. One deficiency of MC method is that it provides only a probabilistic error bound so there is never any guarantee that the expected accuracy is achieved in an actual calculation. Therefore, to increase efficiency and reduce the variance, several techniques of variance reduction sampling were developed. One of these methods is importance sampling, which was mentioned in the previous section. Latin Hypercube sampling (LHS, [45–47], descriptive sampling [48] and HSS [13] were developed with the same aim of variance reduction.

In conventional MC sampling technique, the successive nodes in the sample are independent of each other, which show the measure of randomness in the sample. However, a detailed analysis by [49] showed that for many cases, the true randomness of the samples is not relevant but more importantly the samples should be spread in a uniform manner over the integration domain. Stratified sampling techniques were developed with this idea in mind. LHS is one form of stratified sampling that can yield more precise estimates of the distribution function. In LHS, the range of each uncertain parameter X_i is sub-divided into non-overlapping intervals of equal probability. One value from each interval is selected at random with respect to the probability distribution in the interval. The n values thus obtained for X_1 are paired in a random manner (i.e. equally likely combinations) with n values of X_2 . These n values are then combined with n values of X_3 to form n -triplets, and so on, until n k -tuplets are formed. One variant of this stratified sampling technique is median Latin Hypercube sampling (MLHS) which is similar to descriptive sampling described by Saliby [48]. One drawback of this method is the fact that this stratification scheme is uniform only in one dimension and does not provide uniformity in k -dimensions.

We mentioned that MC method provides us with probabilistic error bounds. However, deterministic error bounds can be obtained if deterministic nodes are used. This leads to the idea of selecting deterministic nodes in such a way to obtain an error bound which is as small as possible. This is the fundamental idea behind quasi-MC methods, where deterministic sequences are used which enhances a certain statistical property without showing significant departure from randomness. A deterministic error bound is obtained, which is smaller than the probabilistic MC error bound and this leads to significant

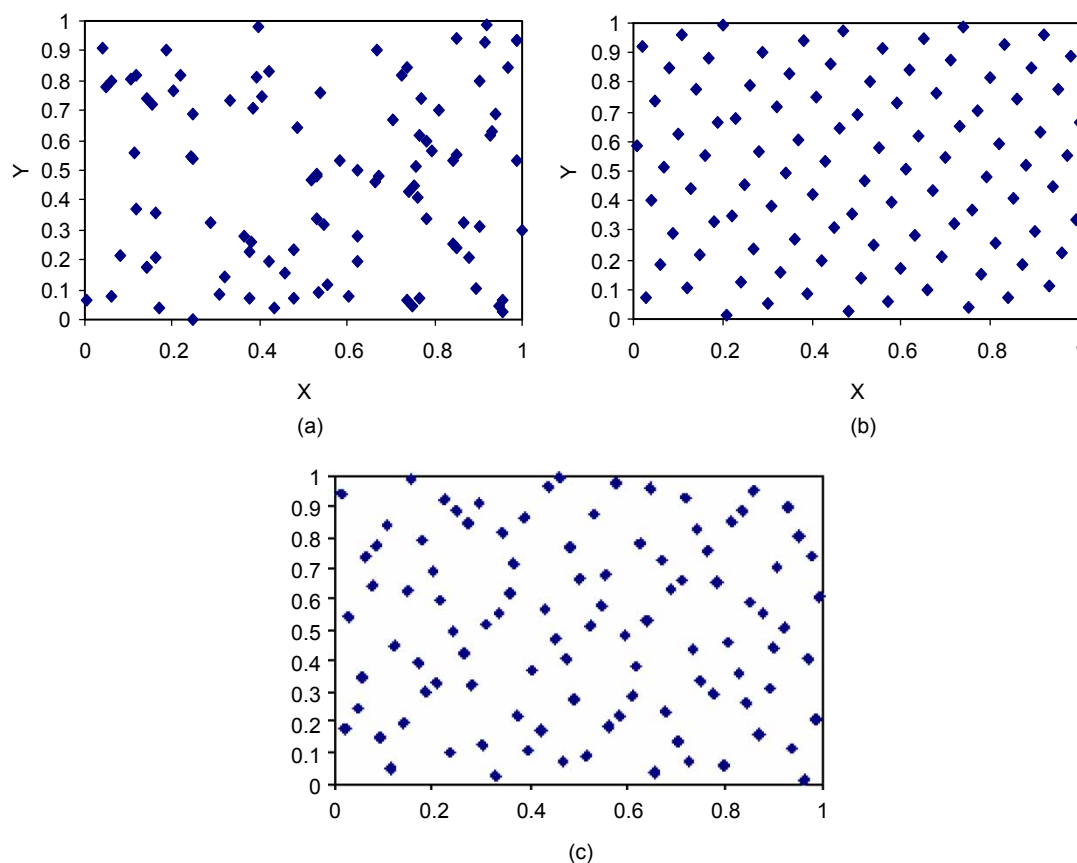


Figure 3. 100 sample points on a unit square using (a) subtract-and-borrow pseudo-random number generator (Marsaglia *et al.*, 1990) [59], (b) Hammersley sequence, (c) shifted Hammersley sequence.

gains in efficiency. Neiderreiter [50] has presented an extensive review of the quasi-MC methods. One example of this is HSS [13] based on Hammersley points. This sampling technique uses an optimal design scheme for placing the n points on a k -dimensional Hypercube, which ensures uniformity properties in multi-dimensions. In contrast, other stratified techniques such as the Latin Hypercube are designed for uniformity along a single dimension and then randomly paired for placement on a k -dimensional cube. One of the main advantages of MC methods is that the number of samples required to obtain a given accuracy of estimates does not scale exponentially with the number of uncertain variables. HSS preserves this property of Monte Carlo. It was found by numerical tests that the HSS technique is at least 3–100 times faster than LHS and MC techniques. Figure 3 shows a comparison of 100 samples of a pseudo-random sequence and HSS in two dimensions. Another important sequence is the shifted Hammersley sequence or the Halton sequence [51], also shown in figure 3, which is very similar to HSS after removing the first variable of Hammersley sequence from the sample. This may be more suitable strategy to use in molecular simulations because instead of generating all the random numbers *a priori*, we can add random numbers to previously generated values. Therefore, in the second case study we use the shifted Hammersley sequence.

We can see from figure 3 that both the Halton sequence and the Hammersley sequence show better uniformity than the pseudo-random number generator which is used in conventional MC simulations. This leads to reduction in error and significant gains in efficiency. The aim of this work is to use these properties of quasi-random sequences to improve the efficiency of MC simulations for the predictions of equations of state and octanol–water partition coefficients which are crucial to environmentally benign and efficient product and process design.

4. Monte Carlo simulations revisited—case studies

As mentioned in Section 3, the aim of quasi-MC methods is to replace the pseudo-random sequences used in conventional MC simulations with quasi-random sequences which show better uniformity properties. In this work, we are applying this idea to MC molecular simulations. Two case studies are considered for this purpose. The first case study involves EOS predictions for a Lennard Jones fluid and the second case study is the prediction of octanol–water partition coefficients by Gibbs free energies of transfer.

4.1 Case study I: equation of state predictions

The case study presented here performs MC simulations of a Lennard Jones fluid to represent a simple system, including

mono-atomic fluids such as argon. The simulations are performed in an NVT ensemble: a system of a fixed number of particles (N) in a given volume (V) at a temperature (T). For this case study, between 100 and 1000 LJ molecules are used in the simulation cell. Periodic boundary conditions are imposed across adjacent simulation cells so as to simulate a bulk fluid. All the quantities in this study are expressed in reduced units, for which more explanations can be found in the literature [17]. The MC simulations are performed at the temperature of $T = 2.0$ in reduced units, much above the critical temperature $T_c = 1.316$ for the LJ fluid. These simulations are performed to predict the EOS of the fluid. To calculate this EOS, pressure values are predicted for different densities of the LJ fluid ranging from $\rho = 0.1$ to 0.9 in reduced units. Pressure for a given density is calculated using the intermolecular pair virial function given in equation (4).

A simple truncated (not shifted) potential $u(r)$ is used as the potential model, where the cutoff radius r_c is set to half the box length L as given in equation (1). The contribution of the particles beyond this cutoff is estimated with the appropriate tail corrections (equations (2) and (3)). These corrections are then applied to the simulation results in order to approximate the properties of the full LJ fluid.

The pressure predictions for the various densities of the LJ fluid, using the above expressions give the EOS (P vs. ρ curve) for the LJ fluid under the given operating conditions. Published results [52] of the EOS calculations for the LJ fluid using pseudo-random numbers from the MC sampling, in the MC and MD simulations, are used as a benchmark for comparison in all the simulation experiments below.

The pseudo-random samples in conventional MC algorithm are replaced with quasi-random HSS samples for 3-dimensional particle moves. The details of this approach are given in figure 4. The crucial point to emphasize here is the choice of random numbers to replace with HSS samples. For example the first random number which is used to choose a random molecule Ran_{MC1} is a discrete random number. Therefore, it is preferable to use a discrete random number generator for this variable. On the other hand, the three random numbers used to perturb molecule positions are linked to the integral and hence

uniformity property in this 3-dimensional space can be exploited to provide maximum gain. This is where the HSS samples generated *a priori* are used instead of pseudo-random samples. The final variable Ran_{MC5} is used for biasing in importance sampling and it should be independent of the previous moves. Hence, pseudo-random numbers are used for this variable.

The equations of state for a 100 and 1000 particle Lennard Jones system are predicted using the NVT ensemble. The HSS samples have increased the efficiency of the MC algorithm, reaching equilibration faster. This is confirmed by comparing the radial distribution function plots of HSS approach and conventional MC algorithm. These plots are shown in figure 6.

The radial distribution function is an indicator of whether the system has reached equilibration or not. This function gives the probability of finding a pair of atoms a distance r apart, relative to the probability expected for an ideal gas at the same density. By definition $g(r) = 1$ is an ideal gas. Thus to test the equilibration, the $g(r)$ curve at the end of equilibration phase can be plotted and compared with the $g(r)$ curve for the standard LJ fluid (figure 5). For a well-equilibrated molecular system configuration the $g(r)$ plots should converge to those of the LJ fluid.

In this figure it can be seen that, with the HSS approach equilibration is reached in 50 cycles, whereas with the conventional MC algorithm using pseudo-random numbers, equilibration takes 200 cycles. From figure 6, we can conclude that for a 100 particle system, the efficiency in improvement is about 3-fold during the equilibration phase of the algorithm.

In figure 7, the EOS comparisons for 100 particle system are presented and these are compared to the benchmark results obtained by Johnson *et al.* [52]. The results obtained using HSS show significant improvement as far as accuracy is concerned and appear to converge much faster.

4.2 Case study II: prediction of octanol–water partition coefficients

In this case study the effect of HSS samples in increasing the efficiency of molecular simulations for predicting

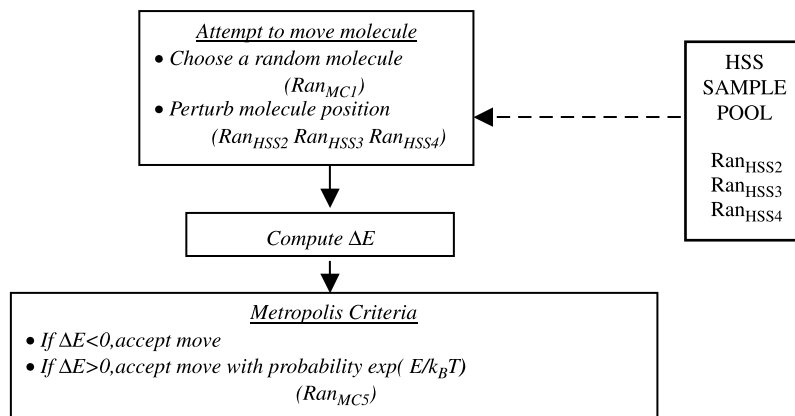


Figure 4. Particle displacement move with the new HSS approach.

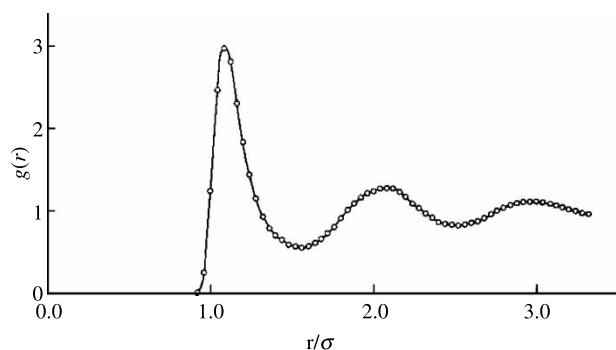


Figure 5. $g(r)$ plot for the LJ fluid close to the triple point ($T = 0.71$, $r = 0.844$).

octanol–water partition coefficients are presented. For this purpose an algorithm similar to what has been proposed by Chen and Siepmann [27] is used, as discussed in Section 2. This algorithm has been implemented using the Towhee MC molecular simulations program (<http://towhee.sourceforge.net>). The flowchart of this algorithm is given in figure 8.

As can be seen from the figure, the octanol–water partition coefficients are predicted by a configurational bias, Gibbs ensemble technique at constant NPT, where five types of MC moves are performed. These moves are carried out for around 10,000 MC cycles, where in one cycle the number of

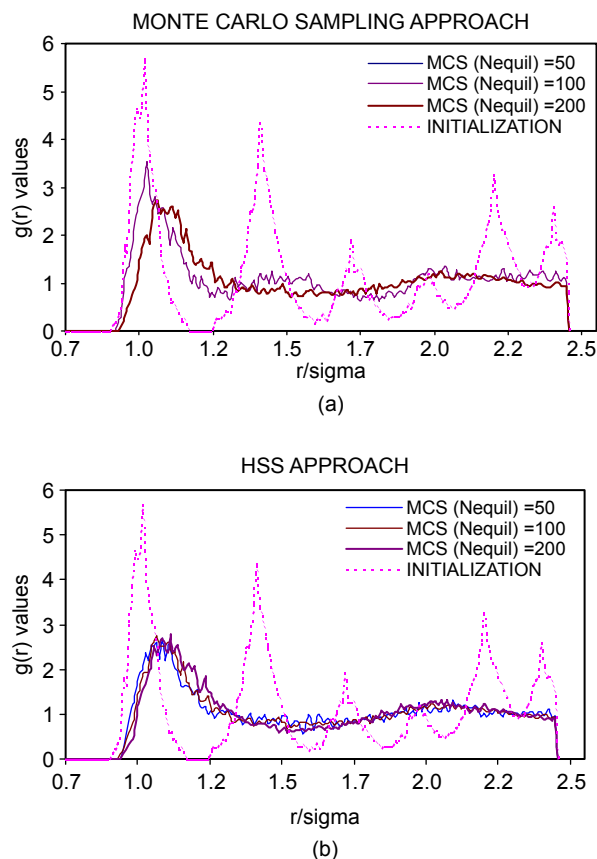


Figure 6. $g(r)$ plot for the (a) conventional Monte Carlo sampling and (b) HSS approach at the initialization of the equilibration and production phases (varying MCS (N_{equil})).

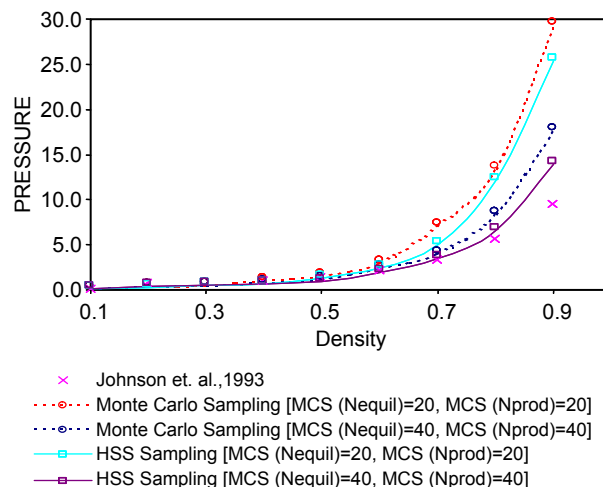


Figure 7. Equation of state predictions—Monte Carlo Sampling approach vs. the HSS approach to MC simulations.

moves performed is equal to the number of molecules (N) in the system. These moves are performed in certain proportions. For example, at every step the probability of performing a volume move is approximately $1/N$. The probability of performing each particle move is equally divided ($(1 - 1/N)/4$). In each of these moves random numbers are required for various purposes.

The overview of each move and the random numbers required for various purposes are provided below:

- Volume move (VOLNPT):** A volume move is defined as the volume change on a single box and the move is accepted based on the energy change, number density, and the specified external pressure. Three random numbers are needed for this move:
 - Select a box at random to change the volume for the box (1 random number)
 - Calculate new volume (1 random number)
 - Acceptance criteria (1 random number)
- Configurational-bias swap move (CBSWAP):** This move performs a 2-box configurational-bias molecule transfer. About $4 + 3$ random numbers are needed for this move:
 - Choose a pair of boxes according to the probabilities (1 random number)
 - Select the primary molecule type (1 random number)
 - Select one choice at random with a bias toward favorable energies (1 random number)
 - Acceptance criteria (1 random number)
 - Select a position for the center of mass uniformly from the entire insertion box—call UNIFORMBOX (3 random numbers)
- Configurational-bias regrowth move (CBREGROW):** This move performs a configurational-bias molecule regrowth move. Three random numbers are needed for this move:
 - Select a molecule type at random (1 random number)
 - Select a proper box (1 random number)
 - Acceptance criteria (1 random number)

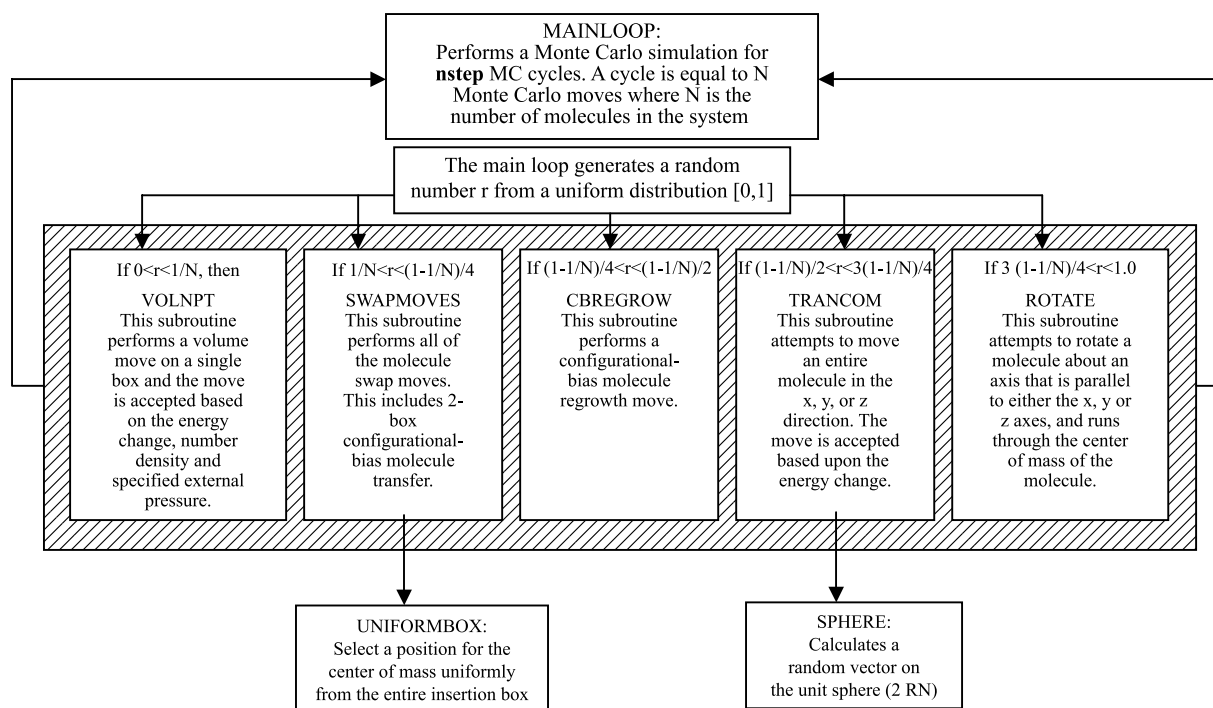


Figure 8. Flowchart of the algorithm for configurational bias Gibbs ensemble simulations for predicting Gibbs free energies of transfer.

- *Rotational move (ROTATE)*: In this move an attempt is made to rotate a molecule about an axis that is parallel to either the x , y or z axes, and runs through the center of mass of the molecule. Six random numbers are needed for this move:
 - Select a molecule type at random (1 random number)
 - Select a proper box (1 random number)
 - Choose a random angular displacement over the three axes (3 random numbers)
 - Acceptance criteria (1 random number)
- *Translational move (TRANCOM)*: In this move an attempt is made to move an entire molecule in the x , y or z direction. The move is accepted based upon the energy change. About $3 + 2$ random numbers are needed for this move.
 - Select a molecule type at random (1 random number)
 - Select a proper box (1 random number)
 - Calculate a random vector on the unit sphere—call SPHERE (2 random numbers)
 - Acceptance criteria (1 random number)

The crucial point here is which random numbers to replace with HSS samples so as to preserve the k -dimensional uniformity to obtain the maximum gain. Therefore, in this case study, only some of the random numbers are replaced by HSS samples, which are summarized in figure 9. In this work, Halton sequence is used, where the first variable of the Hammersley sequence is not taken into account, when replacing the random samples.

From figure 9, it can be seen that choosing an angular displacement over three axes (3-dimensional move), requires three random numbers and for this purpose we

replace the MC samples with 3D HSS samples in this subroutine. Similarly, to exploit the multidimensional uniformity properties of HSS, we replace the two random numbers required to calculate a random vector on a unit sphere with 2D HSS samples. Furthermore, in all of the moves except VOLNPT, a molecule type is chosen to perform the selected MC move. Therefore, in moves SWAPMOVES, CBREGROW, ROTATE and TRANCOM, we replace the four random numbers with four HSS samples. The HSS samples are generated *a priori* and replaced by random samples during the simulation.

The Gibbs free energies of transfer and the partitioning of 4 alcohol solutes (methanol, ethanol, n -propanol, n -butanol) between water and pure 1-octanol are computed from number densities using molecular simulations. The TIP4P and OPLS united atom force fields are used to describe the interactions of water and alcohols, respectively [53–56]. Simulations are carried out at the isobaric ensemble where the temperature $T = 298$ K and pressure $p = 101.3$ kPa. Two different simulations are performed. One of them is for helium/(dry)1-octanol and the other one helium/water system. The total number of particles for these two systems varies between 75 and 126 molecules. The helium/(dry)1-octanol system consisted of helium (50), n -butanol (1), n -propanol (1), ethanol (1), methanol (1) and 1-octanol (20) with a total of 74 molecules. The helium/water system consisted of helium (50), n -butanol (1), n -propanol (1), ethanol (1), methanol (1) and water (72) with a total of 126 molecules.

In order to analyze which method reaches equilibration faster, the radial distribution functions $g(r)$ plots should be presented. The radial distribution function is expected

MOVES	VOLUME MOVES	CB - SWAP MOVES	CB - REGROWTH MOVES	TRANSLATIONAL MOVES	ROTATIONAL MOVES
Description	Volume change on a single box	Configurational bias 2 box molecule transfer	Configurational bias regrowth move	Move an entire molecule in x,y,z direction	Rotates a molecule about an axes parallel to x, y or z
Select a molecule type	-	HSS1	HSS2	HSS3	HSS4
Select a proper box	R	R	R	R	R
Calculate new volume	R	-	-	-	-
Choose a random angular displacement	-	-	-	-	HSS1, HSS2, HSS3 for 3 axes
Select a center of mass uniformly	-	R1, R2, R3 for each molecule	-	-	-
Calculate random vector on a unit sphere	-	-	-	HSS1, HSS2	-
Acceptance criteria	R	R	R	R	R
Total random numbers needed	3 random numbers	3+3 random numbers	3 random numbers	3+2 random numbers	3+3 random numbers

4-D HSS
2-D HSS
3-D HSS

Figure 9. The summary of random numbers replaced with HSS samples.

to have an initial larger peak and it should be fluctuating around 1.0 after the first peak similar to the standard LJ fluid (figure 5).

In figure 10, the radial distribution function plots are presented for the helium/1-octanol system obtained as a result of using the HSS approach and conventional MC sampling using pseudo-random numbers. In figure 10(a) the $g(r)$ plot at the end of 2000 cycles of MC is shown. If we compare figure 10(b) and (c), where 5000 cycles of HSS and 8000 cycles of MC method are shown, respectively, we can see that they are very similar. For both of these systems, the initial peaks are lower than that of 2000 cycles and also the subsequent peaks are closer to 1.0, which shows that the system is reaching equilibration. The fact that these plots are very similar shows that, it takes fewer amounts of cycles to reach a similar state during the simulation when HSS samples are used. In this case, we have an approximately 37.5% gain in efficiency.

A similar behavior is also observed for the helium/1-water system, when we look at the radial distribution function plots shown in figure 11 for this system. In figure 11(a), the $g(r)$ plot at the end of 5000 cycles of conventional MC method is shown. In figure 11(b) and (c), the $g(r)$ plots at the end of 5000 cycles of HSS and 10,000 cycles of MC method are shown, respectively. In this case, again the initial peaks for both 5000 cycles of HSS and 10,000 cycles of MC are lower than 5000 cycles of MC, which shows that the system is reaching equilibration. Also the subsequent peaks are

closer to 1.0 and the fluctuations are less. We can observe that these two plots are also very similar. This means that again, with the HSS approach to reach a similar state in simulation we need fewer amounts of cycles during the simulations. This results in an efficiency improvement of about 50% for this case.

To compare the two methods, the property calculations are also analyzed. In tables I and II, the initial results for the calculated ΔG values are presented and the values found using the HSS approach and conventional MC simulations (where pseudo-random numbers are used) are shown, respectively, for the water phase (helium/water system) and octanol phase (helium/(dry)1-octanol). For both of these systems, the results found using the HSS approach are closer to the experimental data. In figure 12, we can see how the percent average error in Gibbs free energies of transfer changes with respect to number of cycles for all of the alcohol solutes. The average error is calculated by comparing the experimental value of Gibbs free energies of transfer with the value calculated using molecular simulations for each solute and the average value of the error is taken.

If we look at the percentage error compared with experimental data for the Gibbs free energies of transfer of each of the four alcohol solutes, for the helium/water system, which is shown in figure 12, we can see that after 8000 cycles, the percentage error for four alcohol solutes falls below 10% with the HSS approach. However, for the conventional MC method, the error is still higher than 10%

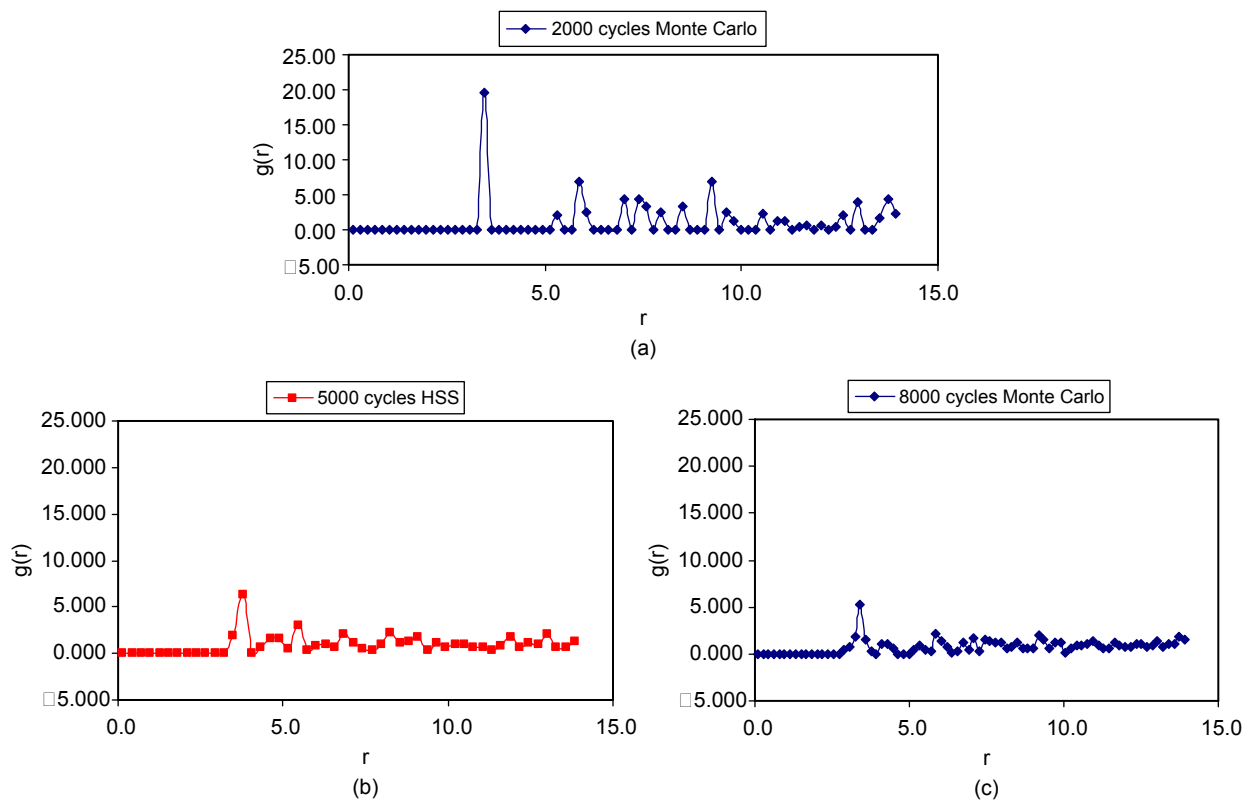


Figure 10. The radial distribution function, $g(r)$ plots for the helium/1-octanol system at the end of: (a) 2000 cycles of conventional MC simulations, (b) 5000 cycles with the HSS approach and (c) 8000 cycles of conventional MC simulations.

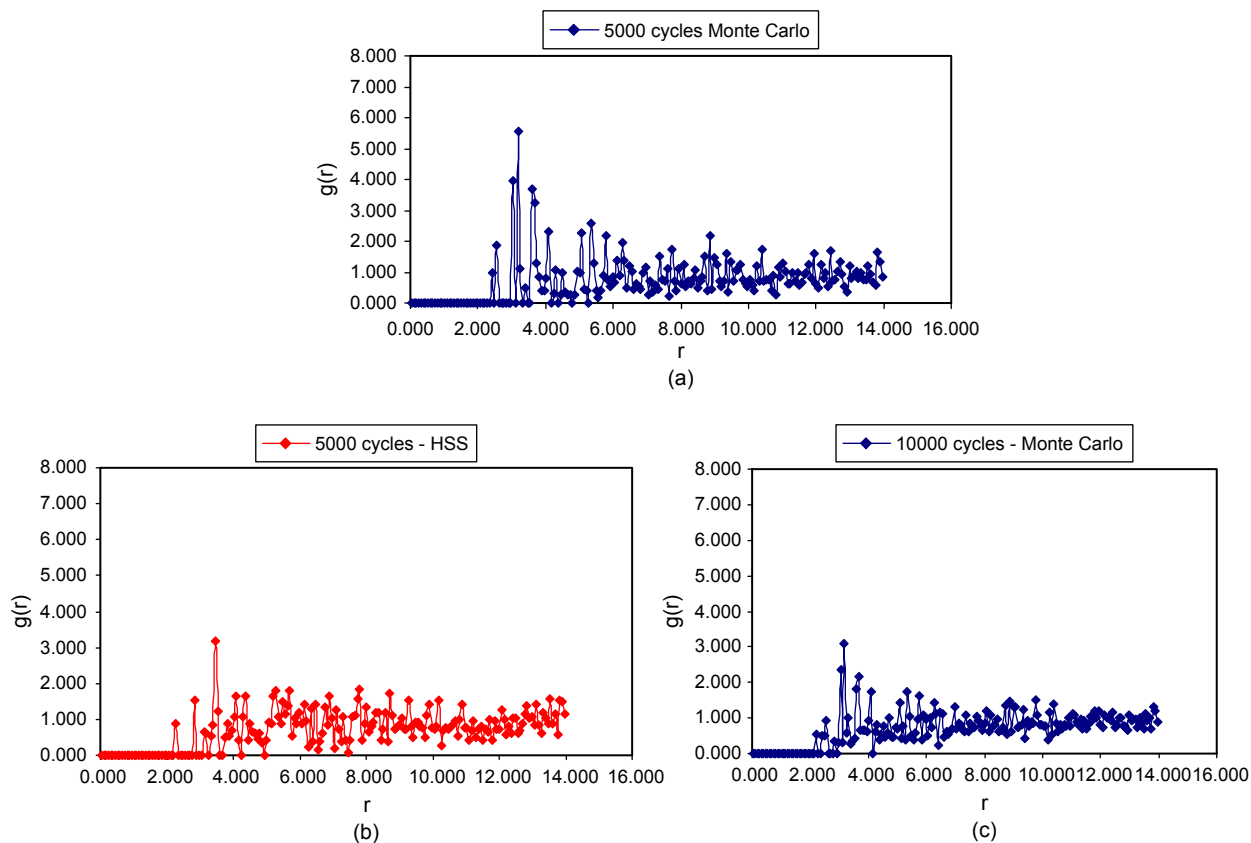


Figure 11. The radial distribution function, $g(r)$ plots for the helium/water system at the end of: (a) 5000 cycles of conventional MC simulations, (b) 5000 cycles with the HSS approach and (c) 10,000 cycles of conventional MC simulations.

Table 1. Gibbs free energies of transfer (in kJ/mol) for 126 particle two-phase system of helium/water at $T = 298$ K and $p = 101.3$ kPa.

Method	Number of cycles	$\Delta G/w$			
		Methanol	Ethanol	Propanol	Butanol
HSS approach	10,000	-19.55	-20.64	-21.26	-20.41
Monte Carlo	10,000	-22.31	-17.12	-17.61	-21.93
Experimental data		-21.19	-20.95	-20.36	-19.86

Table 2. Gibbs free energies of transfer (in kJ/mol) for 75 particle two-phase system of helium/1-octanol at $T = 298$ K and $p = 101.3$ kPa.

Method	Number of cycles	$\Delta G/o$			
		Methanol	Ethanol	Propanol	Butanol
HSS approach	10,000	-14.57	-16.82	-19.51	-18.72
Monte Carlo	10,000	-16.39	-21.47	-16.55	-20.65
Experimental data		-16.17	-18.22	-20.98	-23.88

for ethanol and *n*-propanol (figure 12(b)). This shows that using the HSS approach we can reduce the error in property calculations.

The same situation is also observed for the helium/octanol system which is shown in figure 14. In figure 13(a) the percentage error in Gibbs free energies of transfer are shown for each of the four alcohol solutes when HSS approach is used. The percentage error in Gibbs free energies of transfer falls below 30% around 3000 cycles

for all of the four alcohol solutes. For the conventional MC method shown in figure 13(b), the error is still higher than 30% for three alcohol solutes at 3000 cycles. Also when we compare the results at 8000 cycles, for three alcohol solutes the percentage error is below 10–12% with the HSS approach and with the conventional MC method the percentage error in only one alcohol solute is below this value.

The main aim of this case study is to calculate octanol–water partition coefficients. Therefore, we have computed the

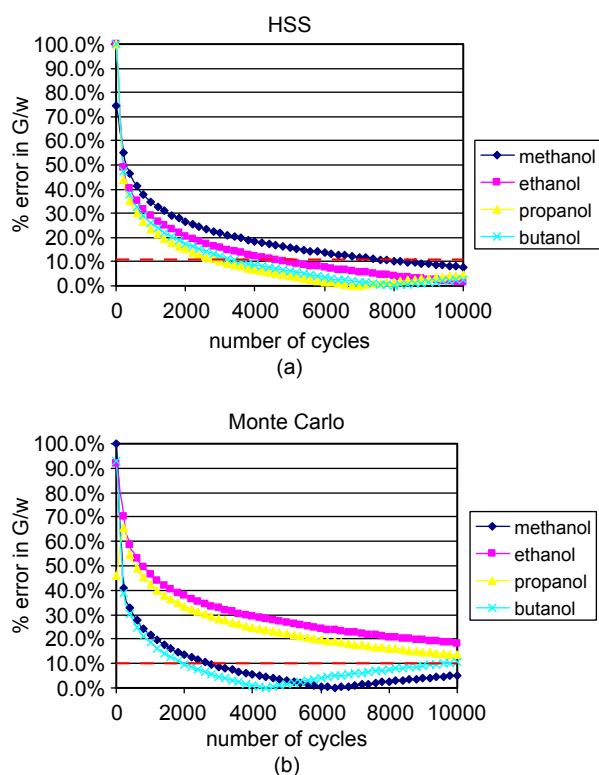


Figure 12. (a) The percentage error between experimental data and calculated values for Gibbs free energy of transfer for each the four alcohol solutes for helium/water system for using the HSS approach, (b) The average percentage error between experimental data and calculated values for Gibbs free energy of transfer for each of the four alcohol solutes for helium/water system using the conventional MC method.

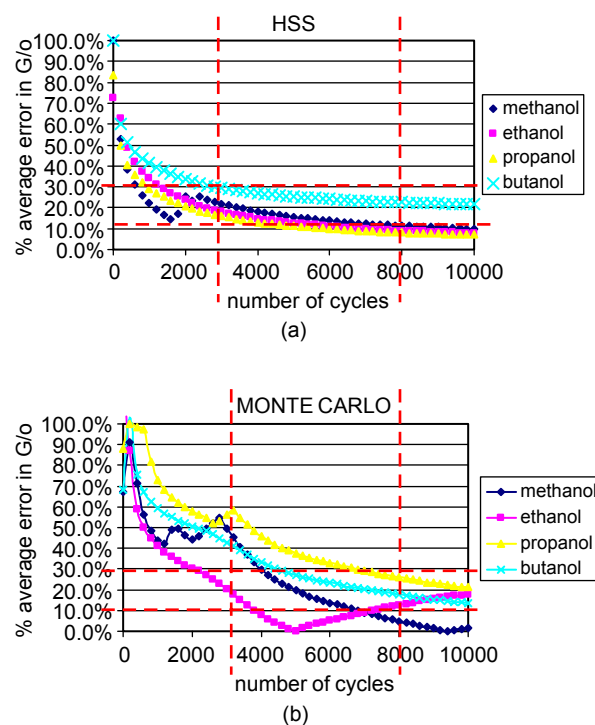


Figure 13. (a) The percentage error between experimental data and calculated values for Gibbs free energy of transfer for each the four alcohol solutes for helium/1-octanol system for using the HSS approach, (b) The average percentage error between experimental data and calculated values for Gibbs free energy of transfer for each of the four alcohol solutes for helium/1-octanol system using the conventional MC method.

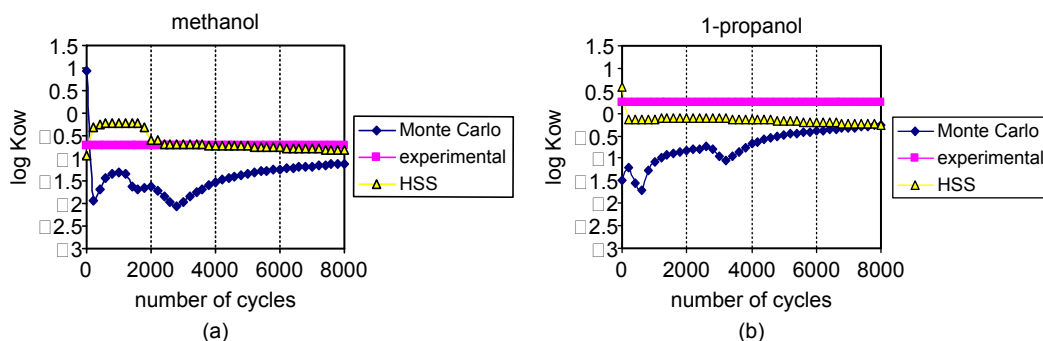


Figure 14. (a): The octanol–water partition coefficient computed using HSS and Monte Carlo technique with respect to number of cycles for methanol, (b) The octanol–water partition coefficient computed using HSS and Monte Carlo technique with respect to number of cycles for propanol.

values of octanol–water partition coefficients $\log K_{ow}$ for the alcohol solutes and the results are shown for methanol and 1-propanol in figure 14(a), (b), respectively. These values are computed from the equation $\ln K_{ow} = G_{ow}/-RT$, where G_{ow} is the difference between the Gibbs free energies of transfer computed from helium/octanol system and the helium/water system. Similar results are also obtained for 1-butanol and ethanol. From this figure, it can be seen that the $\log K_{ow}$ value computed by HSS converges after about 2000 cycles and it does not fluctuate significantly after this point. However, when the MC method is used the value of $\log K_{ow}$ predicted reaches a level of accuracy comparable to HSS only around 8000 cycles. It can be seen that the efficiency improvement is 3-fold. Further, at 2000 cycles where HSS converges, the results are much closer to experimental value.

5. Summary and conclusions

Octanol–water partition coefficient (K_{ow}) is a useful parameter for predicting ecological toxicity (LC_{50} and LD_{50}), bio-accumulation of chemicals inside the body (bio-concentration factor) and provides a basis for QSAR models, which relate the molecular structure of chemicals to their biological activity by a mathematical model. On the other hand, equations of state allow us to calculate the phase equilibria and thermodynamic properties of chemicals. The octanol–water partition coefficient and equations of state provide us with tools for efficient and environmentally benign product and process design and they need to be accurately correlated and predicted. Molecular simulations are useful for predicting these parameters but the computational intensity of this technique hinders its use.

In this work, we presented an efficient approach for increasing the efficiency of molecular simulations. The approach is based on replacing some of the pseudo-random numbers used in molecular simulations by quasi-random numbers generated by the HSS technique. The most important property of this sampling technique is its k -dimensional uniformity property. While replacing the pseudo-random numbers with these quasi-random numbers, we maintain the k -dimensional uniformity property in order to maximize the gains in efficiency.

Two case studies presented in this paper, one involving the calculation of EOS for a Lennard Jones fluid and the second one for the prediction of Gibbs free energies of transfer to be used in octanol–water partition coefficient calculations show the advantages of using HSS samples showing this k -dimensional uniformity property. For the EOS predictions, the efficiency improvement in the equilibration phase is about 3-fold. Similar results are obtained for the octanol–water partition coefficient predictions. Additionally, the properties predicted using the HSS approach are much closer to experimental data.

In summary, it has been shown that the new approach reduces the computational time for molecular simulations and improves accuracy significantly. Furthermore, this technique deals with the random numbers in MC simulations directly and hence can be used in conjunction with any biased MC sampling strategies. Therefore, the applicability of this approach is not restricted to specific systems or specific property predictions.

Finally, we would like to comment that any random or quasi-random sequence such as the one used in this work, should be tested for performance before its use for property predictions.

Acknowledgements

The funding provided by National Science Foundation through the grant #CTS 0353088 is gratefully acknowledged. We also thank Dr Manish Tayal and Dr Kemal Sahin for the preliminary work related to equation of state calculations.

References

- [1] U.M. Diwekar. Greener by design. *Environ. Sci. Technol.*, **37**, 5432 (2003).
- [2] S.E. DeBolt, P.A. Kollman. Investigation of structure, dynamics and solvation in 1-octanol and its water-saturated solution: molecular dynamics and free-energy perturbation studies. *J. Am. Chem. Soc.*, **117**, 5316 (1995).
- [3] J. Gmehling, U. Onken. *Vapor Liquid Equilibrium Data Collection*, Dechema, Flushing, NY Frankfurt/Main (1974).
- [4] K.-J. Kim, U.M. Diwekar. Hammersley stochastic annealing: efficiency improvement for combinatorial optimization under uncertainty *IIE Trans. IIE Trans.*, **34**(9), 761 (2002a).
- [5] K.-J. Kim, U.M. Diwekar. Efficient combinatorial optimization under uncertainty. 2. Application to stochastic solvent selection. *Ind. Eng. Chem. Res.*, **41**, 1285 (2002b).

- [6] D. Crosby. *Environmental Toxicology and Chemistry*, Oxford University Press, New York (1998).
- [7] C. Hansch, T. Fujita. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, **86**, 1616 (1964).
- [8] C. Gao, R. Govind, H.H. Tabak. Application of the group contribution method for predicting the toxicity of organic chemicals. *Environ. Toxicol. Chem.*, **11**, 631 (1992).
- [9] G. Wienke, J. Gmehling. Prediction of octanol–water partition coefficients, Henry coefficients, and water solubilities using UNIFAC. *Toxicol. Environ. Chem.*, **65**, 57 (1998).
- [10] S.-T. Lin, S.I. Sandler. Prediction of octanol–water partition coefficients using a group contribution solvation model. *Ind. Eng. Chem. Res.*, **38**, 4081 (1999).
- [11] J. Marrero, R. Gani. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. *Ind. Eng. Chem. Res.*, **41**(25), 6623 (2002).
- [12] I. Linkov, M.R. Ames, E.A.C. Crouch, F.K. Satterstrom. Uncertainty in octanol–water partition coefficient: implications for risk assessment and remedial costs. *Environ. Sci. Technol.*, **39**(18), 6917 (2005).
- [13] J.R. Kalagnanam, U.M. Diwekar. An efficient sampling technique for off-line quality control. *Technometrics*, **39**, 308 (1997).
- [14] U.M. Diwekar, J.R. Kalagnanam. Robust design using an efficient sampling technique. *Comput. Chem. Eng.*, **20**, S389 (1996).
- [15] Y. Fu, U.M. Diwekar. An efficient sampling approach to multi-objective optimization. *Ann. Oper. Res.*, **132**, 109 (2004).
- [16] M.C. Tayal, U.M. Diwekar. Novel sampling approach to optimal molecular design under uncertainty. *AIChE J.*, **47**, 609 (2001).
- [17] D. Frenkel, B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego (1996).
- [18] S.A. Best, K.M. Merz, C.H. Reynolds. Free Energy perturbation study of octanol/water partition coefficients: comparison with continuum GB/SA calculations. *J. Phys. Chem. B*, **103**, 714 (1999).
- [19] A.P. Lyubartsev, S.P. Jacobsson, G. Sundholm, A. Laaksonen. Solubility of organic compounds in water/octanol systems—an expanded ensemble molecular dynamics simulation study of log *P* parameters. *J. Phys. Chem. B*, **105**, 7775 (2001).
- [20] B. Chen, J.I. Siepmann. Partitioning of alkane and alcohol solutes between water and (dry or wet) 1-octanol. *J. Am. Chem. Soc.*, **122**, 6464 (2000a).
- [21] J.I. Siepmann, D. Frenkel. Configurational bias Monte–Carlo—a new sampling scheme for flexible chains. *Mol. Phys.*, **75**(1), 59 (1992).
- [22] T.J.H. Vlught, M.G. Martin, B. Smit, J.I. Siepmann, R. Krishna. Improving the efficiency of the configurational-bias Monte Carlo algorithm. *Mol. Phys.*, **94**(4), 727 (1998).
- [23] M.G. Martin, J.I. Siepmann. Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. 2. United-atom description of branched alkanes. *J. Phys. Chem. B*, **103**(21), 4508 (1999).
- [24] C.D. Wick, J.I. Siepmann. Self-adapting fixed-end-point configurational-bias Monte Carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions. *Macromolecules*, **33**(19), 7207 (2000).
- [25] Z. Chen, F.A. Escobedo. A Configurational-bias approach for the simulation of inner sections of linear and cyclic molecules. *J. Chem. Phys.*, **113**(24), 11382 (2000).
- [26] C.R.A. Abreu, F.A. Escobedo. A novel configurational-bias Monte Carlo method for lattice polymers: application to molecules with multicyclic architectures. *Macromolecules*, **38**(20), 8532 (2005).
- [27] B. Chen, J.I. Siepmann. A novel Monte Carlo algorithm for simulating strongly associating fluids: applications to water, hydrogen fluoride, and acetic acid. *J. Phys. Chem. B*, **104**, 8725 (2000b).
- [28] Y. Okamoto. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Model.*, **22**(5), 425 (2004).
- [29] A. Mitsutake, Y. Sugita, Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, **60**(2), 96 (2001).
- [30] B.A. Berg, T. Neuhaus. Multicanonical algorithms for 1st order phase-transitions. *Phys. Lett. B*, **267**(2), 249 (1991).
- [31] A.P. Lyubartsev, A.A. Martinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov. New approach to Monte–Carlo calculation of the free-energy—method of expanded ensembles. *J. Chem. Phys.*, **96**, 1776 (1992).
- [32] E. Marinari, G. Parisi. Simulated tempering—a new Monte Carlo scheme. *Europhys. Lett.*, **19**, 451 (1992).
- [33] K. Hukushima, K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.*, **65**, 1604 (1996).
- [34] D.D. Franz, D.L. Freeman, J.D. Doll. Reducing quasi-ergodic behavior in Monte Carlo simulation by J-Walking: applications to atomic clusters. *J. Chem. Phys.*, **93**, 2769 (1990).
- [35] R.H. Zhou, B.J. Berne. Smart walking: a new method for Boltzmann sampling of protein conformations. *J. Chem. Phys.*, **107**(21), 9185 (1997).
- [36] C.H. Mak. Stochastic potential switching algorithm for Monte Carlo simulations of complex systems. *J. Chem. Phys.*, **122**(21), 214110 (2005).
- [37] S.H. Choi, M.S. Jhon. Valley restrained Monte Carlo procedure as a method to improve sampling efficiency. *Mol. Simul.*, **22**(2), 123 (1999).
- [38] P. Attard. A grand canonical simulation technique for dense and confined fluids with application to a Lennard–Jones fluid. *J. Chem. Phys.*, **107**(8), 3230 (1997).
- [39] B. Chen, J.J. Potoff, J.I. Siepmann. Adiabatic nuclear and electronic sampling Monte Carlo simulations in the Gibbs ensemble: application to polarizable force fields for water. *J. Phys. Chem. B*, **104**(10), 2378 (2000).
- [40] M. Strnad, I. Nezbeda. Parallelized sampling of the Gibbs ensemble. *Mol. Phys.*, **98**(22), 1887 (2000).
- [41] S.B. Opps, J. Schofield. Extended state-space Monte Carlo methods. *Phys. Rev. E*, **63**05, 5 (2001).
- [42] D. Frenkel. Speed-up of Monte Carlo simulations by sampling of rejected states. *Proc. Natl. Acad. Sci. USA*, **101**(51), 17571 (2004).
- [43] B. Leblanc, B. Braunschweig, H. Toulhoat, E. Lutton. Improving the sampling efficiency of Monte Carlo molecular simulations: an evolutionary approach. *Mol. Phys.*, **101**(22), 3293 (2003).
- [44] M. Eleftheriou, D. Kim, J.D. Doll, D.L. Freeman. Information theory and the optimization of Monte Carlo simulations. *Chem. Phys. Lett.*, **276**(5–6), 353 (1997).
- [45] M.D. McKay, R.J. Beckman, W.J. Conover. A comparison of three methods of selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2), 239 (1979).
- [46] R.J. Iman, M.J. NUREG/CR-3624, SAND83-2365 Shortencarier. *A FORTRAN77 Program and User's Guide for Generation of Latin Hypercube and Random Samples for Use with Computer Models*, Sandia National Laboratories, (1984).
- [47] R.L. Iman, J.C. Helton. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.*, **8**(1), 71 (1988).
- [48] E. Saliby. Descriptive sampling: a better approach to Monte Carlo simulations. *J. Oper. Res. Soc.*, **41**(12), 1133 (1990).
- [49] D.E. Knuth. *The Art of Computer Programming, Volume 1. Fundamental Algorithms*, Addison-Wesley Publishing Co, Reading, MA (1973).
- [50] H. Neiderreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia (1992).
- [51] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, **2**, 84 (1960).
- [52] J.K. Johnson, J.A. Zolweg, K.E. Gubbins. The Lennard Jones equation of state revisited. *Mol. Phys.*, **78**, 591 (1993).
- [53] W.L. Jorgensen, J.D. Madura, C.J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.*, **106**, 6638 (1984).
- [54] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926 (1983).
- [55] W.L. Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.*, **90**, 1276 (1986).
- [56] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives. Development and testing of OPLS All-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **118**, 11225 (1996).
- [57] J. Sangster. *Octanol–Water Partition Coefficients: Fundamentals and Physical Chemistry*, John Wiley & Sons, Chichester (1997).
- [58] W.L. Jorgensen, J.M. Briggs, M.L. Contreras. Relative partition coefficients for organic solutes from fluid simulations. *J. Phys. Chem.*, **94**(4), 1683–1686 (1990).
- [59] G. Marsaglia, A. Zaman. A new class of random number generators. *Annals of Applied Probability*, **1**, 462–480 (1991).