Molecular Simulations for Environmental Property Predictions: An Efficient Sampling Approach

Saadet Ulas Acikgoz[a,b], *Urmila M. Diwekar*[a,b]
[a] *Vishwamitra Research Institute, Westmont, IL 60559*
[b] *University of Illinois at Chicago, Department of Bioengineering, 851 S Morgan St, Chicago, IL 60607*

ABSTRACT

An efficient sampling technique based on a quasi-random sequence of points that perform significantly better than conventional Monte Carlo sampling is used to predict octanol-water partition coefficients. This technique is called Hammersley Sequence Sampling (HSS), which uses a low-discrepancy sequence of points with deterministic error bounds and multidimensional uniformity properties. The k-dimensional uniformity properties of HSS have been exploited in order to reduce the number of cycles required to reach equilibration and to obtain more accurate property predictions. Configurational-bias Gibbs ensemble technique was used to predict octanol-water partition coefficients from Gibbs free energies of transfer. Preliminary results using HSS and conventional Monte Carlo sampling (pseudo-random numbers) show that the system reaches equilibration faster, which was confirmed by plotting radial distribution functions. Furthermore, the Gibbs free energies of transfer predicted are closer to experimental data when HSS was used.

1. INTRODUCTION

Octanol-water partition coefficient K_{ow} is an important molecular parameter in estimating various environmental and toxicological properties of chemicals. This coefficient have been used in estimation of toxicity parameters such as LC_{50} (lethal concentration that will kill a statistical 50% of a test population) and LD_{50} (lethal dose that will kill a statistical 50% of a test population). It has also been found that the bio-concentration factor (BCF), describing the bioaccumulation of chemicals in an organism has a strong dependence on K_{ow}. In addition, octanol-water partition coefficients provide a basis to a large amount of quantitative-structure activity relationships (QSARs), which have been used to correlate or predict solute properties in biophases (membranes, fat tissue, body fluids).

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 bioaccumulation 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 (Hansch and Fajita, 1964).

Although, experimental data for K_{ow} is available for more than 18000 chemicals (Sangster, 1997), this number is far from the total number of compounds for which data is desirable. Therefore, molecular modeling simulations have been carried out for this system by various researchers. DeBolt and Kollman (1995) investigated the structure, dynamics and solvation in 1-octanol and its water saturated solution by molecular dynamics and free energy perturbation studies. Using this method they have computed the relative partition coefficients and examined the structural aspects of interaction. Similarly, Best et al. (1999) studied the relative 1-octanol/water partition coefficients by a molecular dynamics/free energy perturbation study and compared their results to a generalized Born/surface area octanol continuum solvation model. More recently, Lyubartsev et al. (2001) studied the solubility of organic compounds in water/octanol systems by molecular dynamics simulations using an expanded ensemble technique for drug compounds. Moreover, studies were performed by Chen and Siepmann (2000) where a configurational-bias Monte Carlo (CBMC) simulation was carried out in the Gibbs ensemble to calculate the partitioning of alkane and alcohol solutes in 1-octanol/water. The advantage of this latter work to free energy perturbation studies is that the Gibbs free energy of transfer can be determined directly from the ratio of solute number densities and the number density ratio can be determined very precisely from CBMC/GEMC simulations with small statistical errors.

The success of Monte Carlo and molecular dynamics methods depends on the efficient sampling of the configuration space, to generate states of low energy, enabling the calculation of properties accurately. In Monte Carlo 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 this paper, in order to reduce the computational time and improve the efficiency of molecular simulations, we have used Hammersley Sequence Sampling (HSS) technique (Kalagnanam and Diwekar, 1997). 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 (Kalagnanam and Diwekar, 1997), robust design of distillation columns (Diwekar and Kalagnanam, 1997), solvent selection (Kim and Diwekar, 2002), multi-objective optimal designs for emission reduction (Fu and Diwekar, 2004) and optimal molecular design under uncertainty (Tayal and Diwekar, 2001). Tayal and Diwekar (2000) used 3D HSS samples for efficient evaluation of 3N-dimensional property integral in molecular simulations. Up to 75% computational savings in the equilibrium phase with an overall savings of 37.5% for a 100 Lennard-Jones particle molecular system was obtained.

In this paper, quasi-random points showing k-dimensional uniformity generated by the HSS technique are used instead of pseudo-random points for various Monte Carlo moves of the molecular simulation algorithm for the prediction of octanol-water partition coefficients, to obtain faster convergence and reduce the computational requirements. A configurational-bias Monte Carlo method based on HSS is derived for Gibbs ensemble to improve the efficiency of molecular simulations to calculate K_{ow}. Using this technique we have significantly reduced the number of Monte Carlo cycles for accurate estimation of Gibbs free energies of transfer and improved the efficiency of equilibration.

This paper is divided into 6 parts. Sections 2 talks about the usefulness of octanol-water partition coefficient and various prediction methods for this important parameter. Section 3 describes quasi-random numbers and the HSS technique. In Section 4, the application of this efficient sampling technique to molecular simulations for the prediction of K_{ow} is detailed. In Section 5, we compare the results we have obtained from pseudo-random points and HSS samples and provide a discussion. In Section 6, summary and future work is presented.

2. OCTANOL-WATER PARTITION COEFFICIENT

2.1. Usefulness of the 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 (1964), as shown in Figure 1. 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.

These are the various applications of octanol-water partition coefficients in predicting biological and toxicological properties (Sangster, 1997):

- *Estimation of bioconcentration factor:* K_{ow} is also an important molecular parameter for describing bioaccumulation of chemicals in the environment. Bioaccumulation is the process by which a chemical accumulates in an organism to a higher concentration, than is present in an external source. It has been found that bioconcentration factor (BCF) of a compound depends upon K_{ow}.
-
- *Estimation of toxicity parameters LC_{50} and LD_{50}:* Many direct relationships have been reported between LC_{50} and octanol-water partition coefficients K_{ow} for many compounds. For example a good linear relationship was reported between log LC_{50} values for fish and log K_{ow} for organic compounds including chloro-alkanes, chloro-benzenes, alcohols and ethers.
-
- *Usefulness in quantitative structure-activity relationships:* The quantitative structure-activity relationship (QSAR) relates numerical properties of the molecular structure to its activity by a mathematical model. Octanol-water partition coefficients provide a basis to a large amount of quantitative-structure activity relationships. QSARs have been used to correlate or predict solute properties such as pharmacokinetic characteristics of drug compounds in biophases (membranes, fat tissue, body fluids), and toxicity.

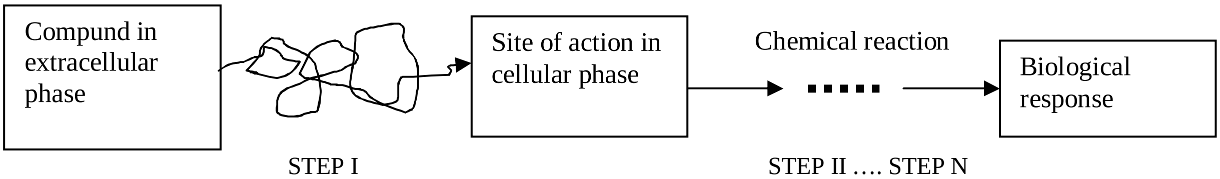


Figure 1: Simplified model for the interaction of a chemical compound and a biological system (Sangster, 1997)

2.2. Prediction of Octanol-Water Partition Coefficients from Molecular Simulations

Calculations of relative octanol-water partition coefficients have been reported using free energy perturbation method with the molecular dynamics and Monte Carlo simulation methods. The computational strategy is as follows (Jorgensen et al., 1990, Best et al., 1999):

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

$$\begin{tikzcd}
A(aq) \arrow[r, "\Delta G_{aq}AB", "2"'] \arrow[d, "\Delta G_{Tr}A"', "1"] & B(aq) \arrow[d, "3"', "\Delta G_{Tr}B"] \\
A(oct) \arrow[r, "4", "\Delta G_{oct}AB"'] & B(oct)
\end{tikzcd}$$

$$\Delta G_{Tr}A = -2.3\ RT \log P_{ow}(A) \tag{1}$$
$$\Delta G_{Tr}B = -2.3\ RT \log P_{ow}(B) \tag{2}$$

In equations 1 and 2, the logarithm of the octanol-water partition coefficient for solutes A and B is 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 3, since we know the fact that free energy is a state function.

$$\Delta G_{Tr}B - \Delta G_{Tr}A = \Delta G_{oct}AB - \Delta G_{aq}AB = \Delta\Delta G_{Tr}AB \tag{3}$$
$$\Delta\Delta G_{Tr}AB = -2.3RT\Delta \log P_{ow} \tag{4}$$

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 4 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 free energy perturbation methods (FEP), molecular dynamics simulations were performed to calculate the relative octanol-water partition coefficients (Best et al. 1999).

Recently, Lyubartsev et al. (2001) 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 Monte Carlo (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.

Another methodology that was used to calculate octanol-water partition coefficients is the configurational-bias Monte Carlo (CBMC) simulations in the Gibbs ensemble (Chen and Siepmann, 2000). 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. As a result, for a given state point the properties of the coexisting phases such as the mutual solubilities of the two solvents and the partitioning of solute molecules can be determined directly from a single simulation.

The main advantages of CBMC/GEMC simulations over free-energy perturbation (FEP) are:

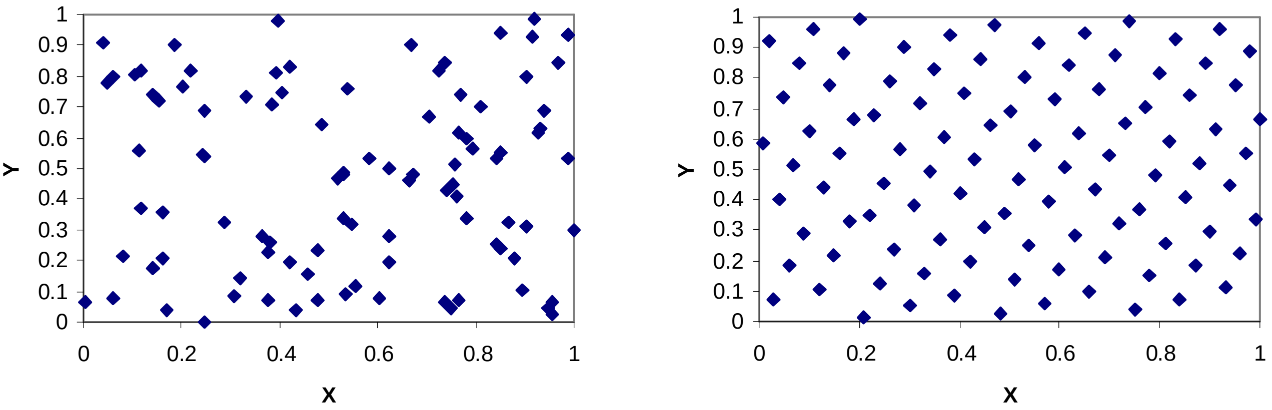
- In both the experiment and CBMC/GEMC simulations Gibbs free energy of transfer is directly determined from the ratio of solute number densities in the two phases (partition constant), while the difference in excess chemical potentials is used in FEP calculations.
- Number density ratio is a mechanical property that can be determined very precisely leading to small statistical errors in ΔG.
- The composition of the two solvent phases do not need to be specified in advance, whereas FEP simulations are performed at a fixed composition that might not correspond to a proper thermodynamic state for the force field used in the calculations.

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 based on quasi-random points. The next section talks about this sampling technique.

3. AN EFFICIENT SAMPLING TECHNIQUE: HAMMERSLEY SEQUENCE SAMPLING

Normal Monte Carlo methods use pseudo random numbers in their calculations. Pseudo random sequences are intended for general use on all classes of problem. However, in quasi-Monte Carlo methods, the basic idea is to use sequences in which a particular statistical property is enhanced without showing significant departure from randomness. Neiderreiter(1992) has presented an extensive review of the quasi-Monte Carlo methods. These methods construct a sequence of points that perform significantly better than Monte Carlo methods with deterministic error bounds.

In this paper, an efficient sampling technique (Hammersley Sequence Sampling) based on Hammersley points, which has been developed by Kalagnanam and Diwekar, (1997) is used. 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 unlike pseudo-random number based Monte Carlo, Latin Hypercube (Iman and Shortencarier, 1984), and its variant, the Median Latin Hypercube (Saliby, 1990) sampling techniques.



(a) (b)

Figure 2: 100 sample points on a unit square using (a) subtract-and-borrow pseudo-random number generator (Marsaglia et al., 1990), (b) Hammersley points

This can be illustrated by a simple experiment. If we place sample points on a unit square using a pseudo random number generator, and Hammersley points we can observe the uniformity properties of HSS. This is illustrated in Figure 2. If the first variable of Hammersley sequence is not taken into account, we obtain a Halton sequence (1960).

Due to its multidimensional uniformity properties, the HSS technique requires fewer samples and has a faster convergence rate varying from 3 to 100 according to the tests performed by Kalagnanam and Diwekar(1997). Next, we will see the effects of this technique on molecular simulations.

4. APPLICATION OF HSS TO MOLECULAR SIMULATIONS

4.1. Results for 100 particle Lennard-Jones system

Tayal and Diwekar(2000) has presented the effect of k-dimensional uniformity properties of HSS samples in efficient sampling of the 3N-dimensional phase space in a Monte Carlo molecular simulations algorithm. The equation of state for a 100 and 1000 particle Lennard Jones system was predicted using the NVT ensemble. The HSS samples have increased the efficiency of the MC algorithm, reaching equilibration faster. This was confirmed by comparing the radial distribution function plots of HSS approach and conventional Monte Carlo algorithm. An example of these plots is shown in Figure 4.

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 more number of cycles. Tayal and Diwekar(2000) have concluded that for a 100 particle system, the computational savings observed are about 75% during the equilibration phase of the algorithm.

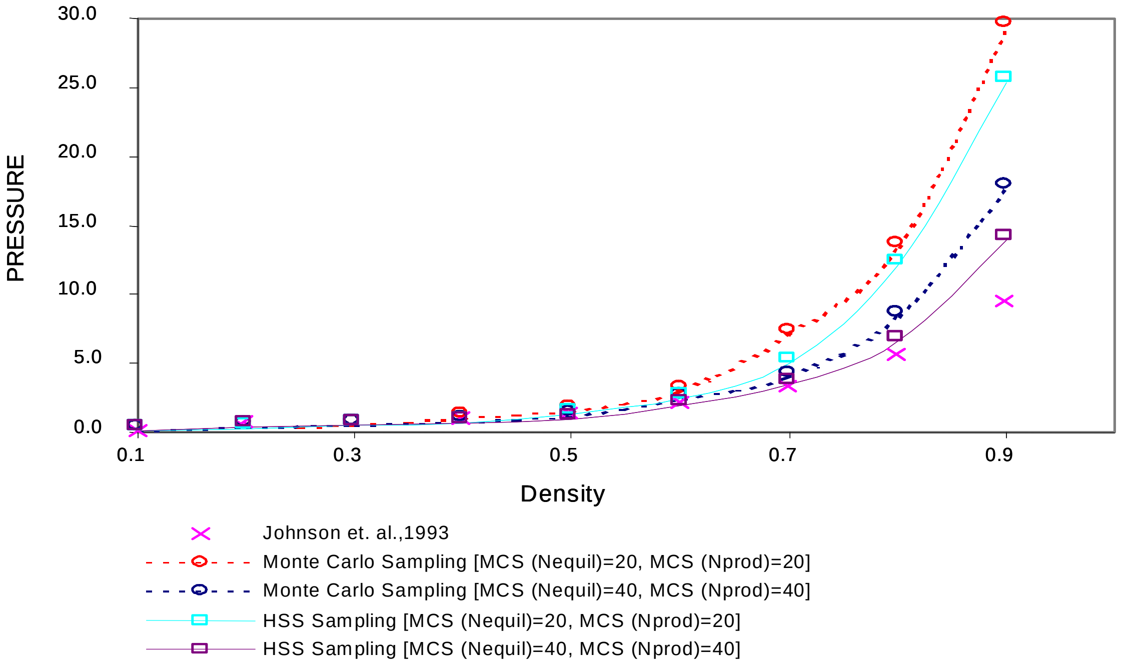
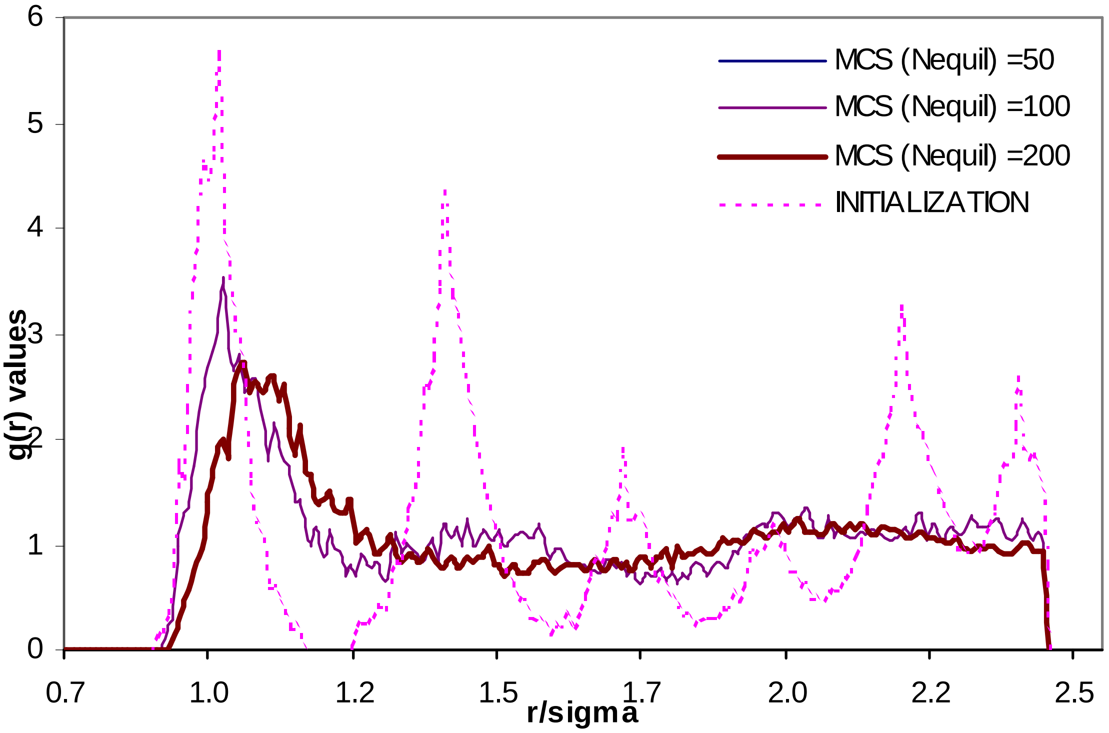


Figure 3: *Equation of state* predictions - Monte Carlo Sampling approach vs. the HSS approach to MC simulations (Tayal and Diwekar, 2000)

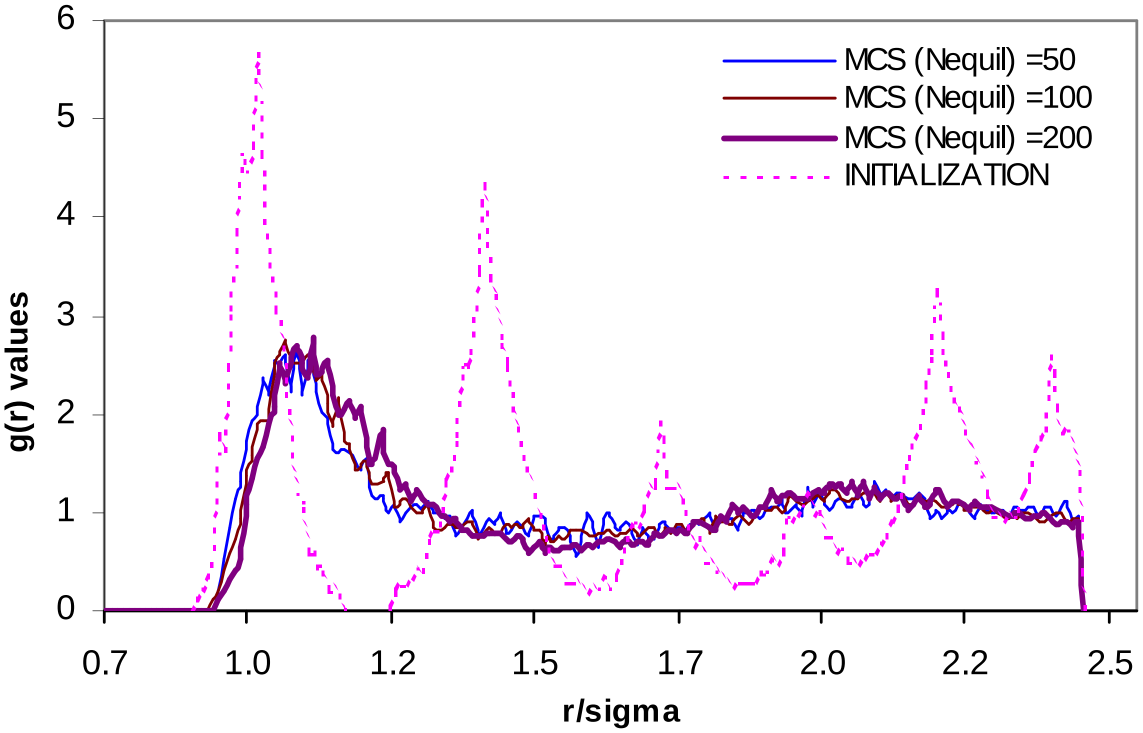
In Figure 3, the equation of state comparisons for 100 particle system are presented and these are compared to the benchmark results obtained by Johnson et al.(1993). The results obtained using HSS show significant improvement and appear to converge much faster.

In the light of these promising results for a Lennard-Jones 100 particle system, the effect of HSS samples in increasing the efficiency of predicting octanol-water partition coefficients are presented in this work. For this purpose an algorithm similar to what has been proposed by Chen and Siepmann (2000) was used, as discussed in Section 2. This algorithm has been implemented using the Towhee Monte Carlo molecular simulations program (http://towhee.sourceforge.net). The flowchart of this algorithm is given in Figure 5.

As can be seen from the figure, the octanol-water partition coefficients are predicted by a configurational bias, Gibbs ensemble technique, where five types of Monte Carlo moves are performed. These moves are carried out for between 100-8000 Monte Carlo cycles, where in one cycle the number of moves performed is equal to the number of molecules (N) in the system. These moves are performed in certain proportions. For example, the volume moves are equal to 0.2% of all the MC moves. 24.8% of these moves are swapmoves and each of configurational bias regrowth moves, rotational moves and translational moves are performed 25% of the time. Each of these moves require random numbers. In order to keep the k-dimensional uniformity of HSS points, we have used different random number generation for each move.



MONTE CARLO SAMPLING APPROACH



HSS APPROACH

Figure 4: g(r) plot for the (a) Conventional Monte Carlo Sampling and (b) HSS approach at the initialization of the Equilibration and Production Phase (varying MCS (N*equil)*) (Tayal and Diwekar, 2000)

The number of random numbers required for each move is given below:

- *VOLNPT (Volume moves):* this part performs a volume change move on a single box and the move is accepted based on the energy change, number density, and the specified external pressure.

  3 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)

- *SWAPMOVES (Swap moves):* this part performs a 2-box configurational-bias molecule transfer

  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)

- *CBREGROW (Configurational-bias regrowth moves):* this part performs a configurational-bias molecule regrowth move.

  3 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)

- *ROTATE (Rotational moves):* this part attempts 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.

  6 random numbers are needed for this part:
  Select a molecule type at random (1 RANDOM NUMBER)
  Select a proper box (1 RANDOM NUMBER)
  Choose a random angular displacement over the 3 axes (3 RANDOM NUMBERS NEEDED)
  Acceptance criteria (1 RANDOM NUMBER)

- *TRANCOM (Translational moves):* this part attempts to move an entire molecule in the x, y, or z direction. The move is accepted based upon the energy change.

  3+2 random numbers are needed for this part.
  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)

It has been found in an earlier work from our group related to simulated annealing (Kim and Diwekar, 2002) that it is important to differentiate where the k-dimensional uniformity is important. Therefore, in our case study, only some of the random numbers was replaced by

HSS samples, which are summarized in Table I. In this work, a Halton sequence was used, where the first variable of the Hammersley sequence was not taken into account, when replacing the random samples.

Table I: The summary of random numbers replaced by HSS samples

MOVE	RANDOM NUMBERS REPLACED	NAME
VOLNPT	Select a box at random to change the volume for the box (1 RANDOM NUMBER)	HSS1
SWAPMOVES	Select the primary molecule type (1 RANDOM NUMBER)	HSS1
CBREGROW	Select the molecule type (1 RANDOM NUMBER)	HSS2
ROTATE	Select the molecule type (1 RANDOM NUMBER) Choose a random angular displacement over the 3 axes (3 RANDOM NUMBERS)	HSS3 HSS1,HSS2,HSS3
TRANCOM	Select the molecule type (1 RANDOM NUMBER)	HSS4
SPHERE(called by TRANCOM)	Calculate a random vector on the unit sphere (2 RANDOM NUMBERS)	HSS1,HSS2

From Table I, it can be seen that choosing an angular displacement over 3 axes (3-dimensional move), requires 3 random numbers and for this purpose we replaced the Monte Carlo samples with 3D HSS samples in this subroutine. Similarly, for sphere to exploit the multidimensional uniformity properties of HSS, we have replaced the 2 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 Monte Carlo move. Therefore, in moves SWAPMOVES, CBREGROW, ROTATE AND TRANCOM, we replaced the four random numbers with four HSS samples.

The HSS samples were generated a priori and replaced by random samples during the simulation. The preliminary results from these simulations are given in the next section.

5. PRELIMINARY RESULTS & DISCUSSION

The Gibbs free energies of transfer and the partitioning of alcohol solutes between water and pure octanol were computed from number densities using molecular simulations. The Towhee Monte Carlo molecular simulations program was used for this purpose. The TIP4P and OPLS united atom force fields were used to describe the interactions of water and alcohols respectively (Jorgensen et al. (1984); Jorgensen et al. (1983);, Jorgensen (1986); Jorgensen and Tirado-Rives (1988)). Simulations were carried out at the isobaric ensemble where the temperature T=298 K and pressure p=101.3 kPa. Two different two-phase systems were considered. One of them was helium/(dry)1-octanol and the other one helium/water system.

In Table II and Table III, the initial results for the calculated ΔG values are presented and the values found using the HSS approach and conventional Monte Carlo 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. For the water phase, the calculated values of ΔG with the HSS method after 8000 cycles, the error is even smaller than the values calculated after 10000 cycles using the conventional Monte Carlo random sampling approach. Note that the results reported in Table II are for a 1510 particle system (this system is composed of helium (600), methane (20), ethane (20), propane(1), n-butane(1), methanol(1), ethanol(1), propanol(1), n-butanol(1), water(864)). In Table III, the results are reported for a 165 particle system (this system is composed of helium(100), methanol(1), ethanol(1), propanol (1), n-butanol(1), 1-octanol(60))

Table II. Gibbs free energies of transfer (in kJ/mol) for 1510 particle two-phase system of helium/water at T=298 K and p=101.3 kPa.

METHOD	**Number of cycles**	**ΔG/w methanol**	**ΔG/w ethanol**	**ΔG/w propanol**	**ΔG/w butanol**
HSS approach	8000	-16.19	-15.10	-18.13	-15.85
Monte Carlo –random	8000	-15.13	-13.32	-17.36	-12.55
Monte Carlo – random	10000	-15.74	-13.92	-17.97	-13.16
experimental data		**-21.19**	**-20.95**	**-20.36**	**-19.86**

Table III. Gibbs free energies of transfer (in kJ/mol) for 164 particle two-phase system of helium/1-octanol at T=298 K and p=101.3 kPa.

METHOD	**Number of cycles**	**ΔG/o methanol**	**ΔG/o ethanol**	**ΔG/o propanol**	**ΔG/o butanol**
HSS approach	3000	-16.18	-16.95	-13.32	-16.74
Monte Carlo –random	3000	-20.36	-24.04	-23.30	-30.17
experimental data		**-16.17**	**-18.22**	**-20.98**	**-23.88**

In Figure 6, the radial distribution function plots are presented for the helium/1-octanol system obtained as a result of using the HSS approach and conventional Monte Carlo sampling using pseudo-random numbers. The simulations were carried out for 10000 cycles and the radial distribution function plots at the end of 1000, 3000 and 10000 cycles are illustrated in this figure. It can be seen clearly that the when we use the HSS approach the system reaches equilibration much faster. This can be observed from the shape of the radial distribution function plots and their closeness to the ideal case where g(r)=1.

The accuracy of the predictions for Gibbs free energies of transfer can further be increased by increasing the number of cycles during the production phase.

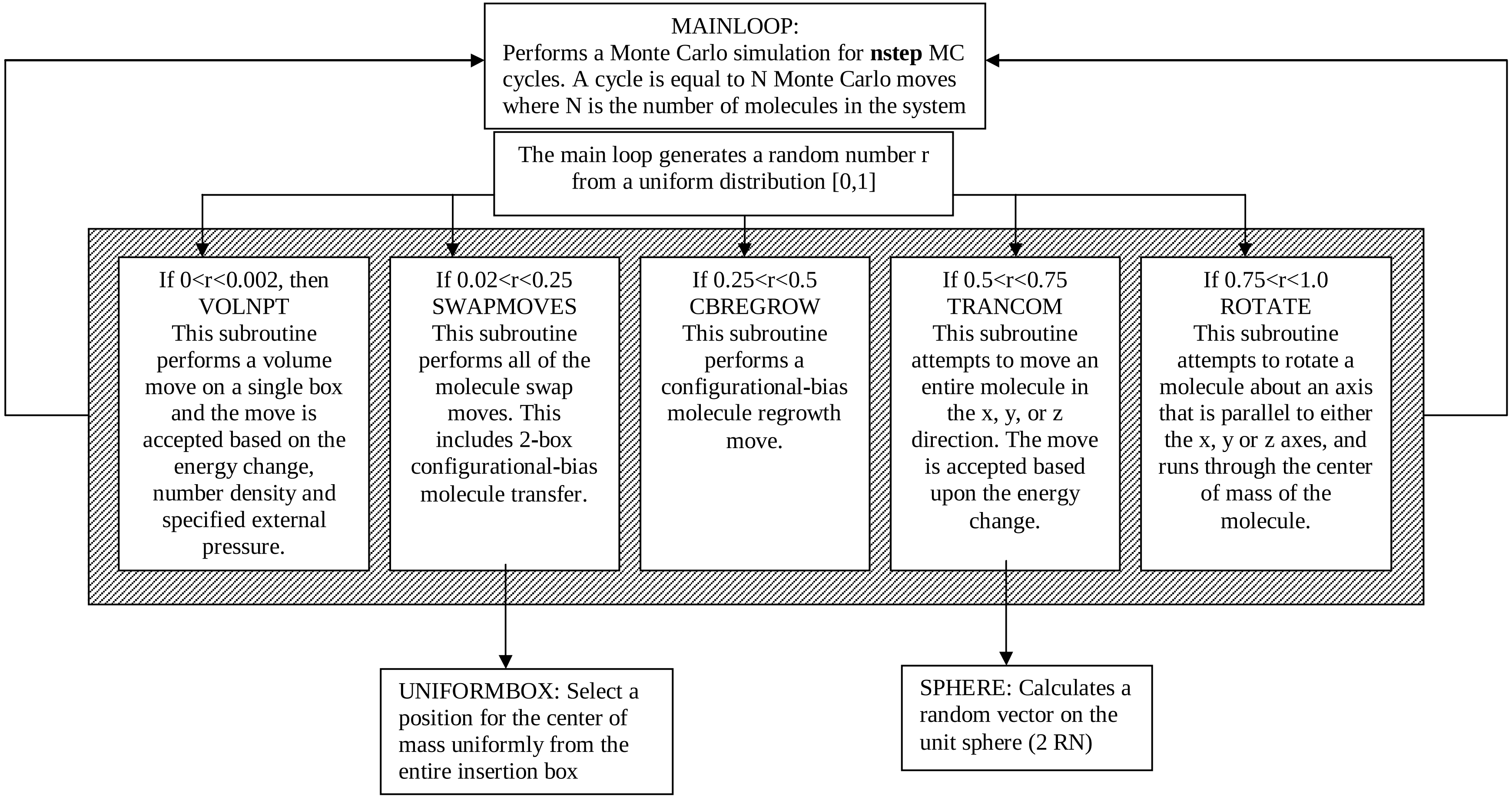


Figure 5: Flowchart of the algorithm for predicting octanol-water partition coefficients

6. SUMMARY AND FUTURE WORK

Octanol-water partition coefficient (K_{ow}) is a useful parameter for predicting ecological toxicity (LC_{50} and LD_{50}), bioaccumulation of chemicals inside the body (bioconcentration factor) and provides a basis for QSAR models, which relate the molecular structure of chemicals to their biological activity by a mathematical model.

In this work, the effect of an efficient sampling technique to increase the efficiency of predicting K_{ow} from Gibbs free energies of transfer by molecular simulations was explored. For this purpose, the random samples at certain points of the simulation program were replaced with samples showing k-dimensional uniformity properties. This sampling technique, called Hammersley Sequence Sampling (HSS) is a technique showing faster convergence properties than conventional Monte Carlo sampling.

The preliminary results obtained with conventional random sampling (MCS) and HSS show that with HSS, predictions for K_{ow} are closer to experimental data and the system reaches equilibration much faster. This was confirmed by radial distribution function plots. As a future work, we are planning to confirm our results with more simulations for the 1-octanol and water phases.

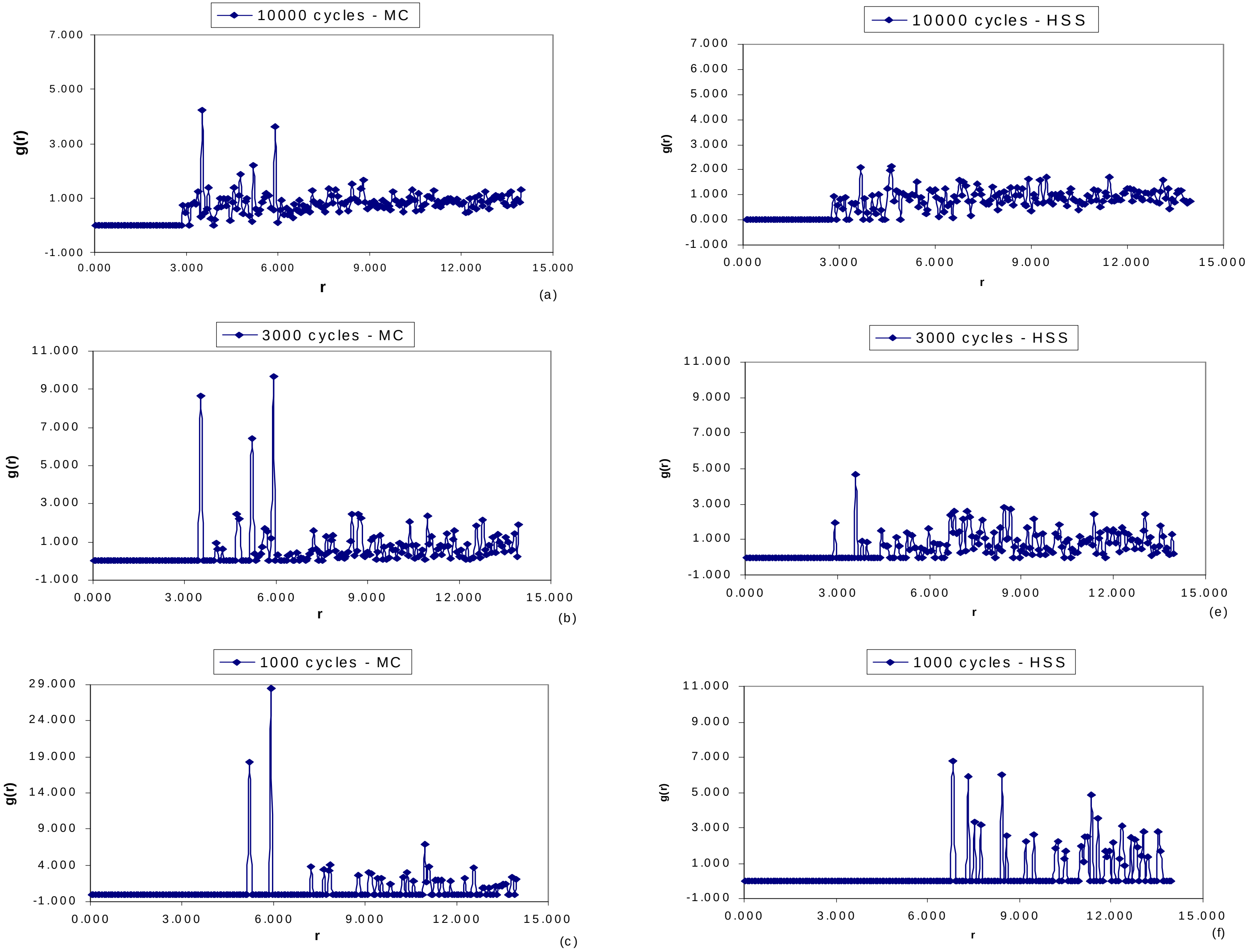


Figure 6: The radial distribution functions(rdf) between two molecules for the 164 particle system of helium/pure 1-octanol. (a) rdf after 10000 cycles with Monte Carlo – random sampling, (b) rdf after 3000 cycles with Monte Carlo – random sampling, (c) rdf after 1000 cycles with Monte Carlo – random sampling, (d) rdf after 10000 cycles with HSS, (e) rdf after 3000 cycles with HSS, (f) rdf after 1000 cycles with HSS

REFERENCES

Best S.A., Merz K.M., Reynolds C.H., Free Energy Perturbation Study Of Octanol/Water Partition Coefficients: Comparison With Continuum GB/SA Calculations, *Journal of Physical Chemistry B* 103: 714-726, 1999

Chen B. & Siepmann J.I., ‘Partitioning of Alkane and Alcohol Solutes between Water and (Dry or Wet) 1-Octanol', Journal of American Chemical Society 122: 6464-6467, 2000.

DeBolt S.E. & Kollman P.A., ‘Investigation of Structure, Dynamics and Solvation in 1-Octanol and its Water-Saturated Solution: Molecular Dynamics and Free-Energy Perturbation Studies', *Journal of American Chemical Society* 117: 5316-5340, 1995

Diwekar U.M. & Kalagnanam J.R., ‘Robust design using an efficient sampling technique'**,** *Computers & Chemical Engineering,* 20: S389-S394, 1996

Fu Y., Diwekar U.M., ‘An efficient sampling approach to multi-objective optimization', *Annals of Operations Research,* 132: 109-134, 2004

Hansch C. & Fujita T., ‘*p*- - Analysis. A Method for the Correlation of Biological Activity and Chemical Structure', *Journal of American Chemical Society* 86: 1616-1626, 1964

Iman, R.J. and Shortencarier, M.J., A FORTRAN77 Program and User's Guide for Generation of Latin Hypercube and Random Samples for Use with Computer Models, NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, N.M., 1984.

Jorgensen W.L., Madura J.D. and Swenson C.J., ‘Optimized Intermolecular Potential Functions for Liquid Hydrocarbons’, J. Am. Chem. Soc., 106:6638-6646, 1984et al. (1984)

Jorgensen W.L., ‘Optimized Intermolecular Potential Functions for Liquid Alcohols’, J. Phys. Chem., 90:1276-1284, 1986(1986)

Jorgensen W.L., Chandrasekhar J., Madura J.D., ‚Comparison of Simple Potential Functions for Simulating Liquid Water’. J.Chem. Phys., 79:926-935, 1983et al. (1983)

Jorgensen W.L., Maxwell D.S., and Tirado-Rives J., ‘Development and Testing of OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids’, J. Am. Chem. Soc., 118:11225-11236, 1996 and Tirado-Rives (1988)

Kalagnanam J.R. & Diwekar U.M., ‘An Efficient Sampling Technique for Off-line Quality Control', *Technometrics* 39: 308 – 319, 1997

Kim K.J. & Diwekar U.M., ‘Efficient Combinatorial Optimization Under Uncertainty. 2. Application to Stochastic Solvent Selection', *Industrial & Engineering Chemistry Research* 41: 1285-1296, 2002

Kim KJ, Diwekar UM ‘Hammersley stochastic annealing: efficiency improvement for combinatorial optimization under uncertainty’ IIE Transactions 34 (9): 761-777, 2002

Lyubartsev A.P., Jacobsson S.P, Sundholm G., & Laaksonen A., ‘Solubility of Organic Compouns in Water/Octanol Systems-An Expanded Ensemble Molecular Dynamics Simulation Study of log P Parameters', *Journal of Physical Chemistry B* 105: 7775-7782, 2001.

Neiderreiter H, ‘Random Number Generation and Quasi-Monte Carlo Methods’, SIAM, Philadelphia, 1992. (1992)

Saliby E. (1990), Descriptive sampling: A better approach to Monte Carlo simulations, J. Operations Research Society,41(12), 1133.

Sansgter J., ‘*Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*', John Wiley & Sons, Chichester, 1997

Tayal M.C., & Diwekar U.M. ‘Novel Sampling Approach to Optimal Molecular Design under Uncertainty', *AIChE Journal,* 47: 609-628, 2001

Tayal M.C., & Diwekar U.M., ‘Improved Molecular Simulations by Efficient Sampling', *AIChE Annual Meeting*, Los Angeles, CA, 2000