Quality of random number generators significantly affects results of Monte Carlo simulations for organic and biological systems

Timothy H. Click[a], **George A. Kaminski**[a,*], and **Aibing B. Liu**[b]

[a] Worcester Polytechnic Institute, Worcester, Massachusetts 01609

[b] Central Michigan University, Mount Pleasant, Michigan 48859

Abstract

We have simulated pure liquid butane, methanol and hydrated alanine polypeptide with the Monte Carlo technique using three kinds of random number generators - the standard Linear Congruential Generator (LCG), a modification of the LCG with additional randomization used in the BOSS software, and the "Mersenne Twister" generator by Matsumoto and Nishimura. While using the latter two random number generators leads to reasonably similar physical features, the LCG produces a significant different results. For the pure fluids, a noticeable expansion occurs. Using the original LCG on butane yields a molecular volume of 171.4 Å^3 per molecule compared to ca. 163.6–163.9 Å^3 for the other two generators, a deviation of about 5%. For methanol, the LCG produces an average volume of 86.3 Å^3 per molecule, which is about 24% higher than the 68.8–70.2 Å^3 obtained with the random number generator in BOSS and the generator by Matsumoto and Nishimura. In case of the hydrated tridecaalanine peptide, the volume and energy tend to be noticeably greater with the LCG than with the BOSS (modified LCG) random number generator. For the simulated hydrated extended conformation of tridecaalanine, the difference in volume reached ca. 87%. The uniformity and periodicity of the generators do not seem to play the crucial role in these phenomena. We conclude that it is important to test a random number generator by modeling a system such as the pure liquid methanol with a well-established force field before routinely employing it in Monte Carlo simulations.

Keywords

Random number generators; Monte Carlo molecular simulations

I. Introduction

Monte Carlo molecular simulations have been an extremely valuable tool in a wide variety of computer modeling applications, from predicting pure liquid densities and heats of vaporization to assessing relative binding energies of protein-ligand complexes.[1] Random number generators (RNG's) are an integral part of Monte Carlo simulations of molecular systems. While mathematical analysis of various randomization techniques is available, there is a lack of literature on the actual effects that a good or poor random numbers generator can have on results of Monte Carlo simulations.

There are two parts of a Monte Carlo simulation in which random numbers are needed. First, an atom or a molecule has to be moved by a random distance and direction. Then the new

---

[*]Author to whom correspondence should be addressed. gkaminski@wpi.edu.

potential energy E(new) and the difference with the old energy E(old) are calculated. In the Metropolis sampling formalism, the following quantity is calculated:

$$A = \exp[-(E(new) - E(old))/RT]$$

(1)

which is then compared to a random number X, where $0 < X < 1$. If $A \geq X$, the attempted move is accepted, and it is rejected otherwise. Generation of the X is the second place where random numbers are used in the Monte Carlo cycle.

The uniformity of the random number distribution is important. On one hand, the nature of Monte Carlo simulations tends to randomize the use of any generator, as it uses random numbers for a number of purposes, including choosing which molecule to move, in which direction, etc. On the other hand, having pseudo-random numbers which tend to be too high or too low will lead to rejecting or accepting a disproportional number of the attempted Monte Carlo steps, respectively. There is an even more important reason for using uniform RNG in Monte Carlo simulations. One of the uses for the generator is to pick a molecule for a random move. If an RNG systematically under-samples a certain region of its domain, there will be molecules which are moved less frequently. This can cause a volume expansion of the system. A similar effect is known to occur when WKC preferential sampling is used with an excessive bias in sampling a part of the system.1g Therefore, it is useful to test uniformity of the random number distribution.

This goal can be achieved with the so called frequency test 2a After generating N pseudo-random numbers, they are ordered from the least to the greatest, in a sorted sequence $\{R_i\}$. Then the following quantities are calculated:

$$d_i = |R_i - X_i|$$

(2)

where $X_i = (i - 0.5)/N$. The greatest of the d_i values is taken as a measure of the deviation from uniformity for the generator. Figure 1 illustrates using this measure:

For a "good" uniform random number generator, the sorted sequence will follow the $y = x$ line closely, since the numbers are generated uniformly on the interval from 0 to 1, and so all parts of it are represented equally.

Another characteristic of a random number generator is the length of the period after which the pseudo-random sequence starts repeating itself. It is obvious that using a random number generator with a longer period should not decrease the quality of the results. At the same time, it is not obvious whether the length of the period itself is crucial for as long as the pseudo-random numbers distribution is uniform enough.

By simulating pure liquid butane and methanol and solvated tridecaalanine with different random number generators, we demonstrated that the choice of a generator can significantly affect the resulting physical properties of the system and showed that this effect does not necessarily correlate with the uniformity of period length of the RNG. The choice of tridecaalanine was made since this it is known to form an alpha-helix in aqueous solution, thus it is small yet biologically relevant molecular system which has been studied both experimentally and computationally. We would like to emphasize it very strongly though that studying the specific folding behavior is in no way a goal of this work. This behavior would depend on the quality of the force field, the length of the simulations and a number of other factors which are not relevant in this specific project. What we are reporting is a significant difference produced by choosing a specific random number generator. The rest of

the paper is organized as follows. Section II introduces the random number generators used in this study and briefly describes the pure liquid simulations as well as the polypeptide modeling carried out with the Monte Carlo technique. Given in Section III are the results and discussion. These are followed by conclusions in Section IV.

II. Methods

A. Random Number Generators

We employed three types of random number generators. The first is the Linear Congruential Generator (LCG). This is a widely used generator. It uses integer constants a, c, and N, and produces a sequence of pseudo-random integer numbers $\{x_i\}$, where the following recursive formula is used:

$$x_{i+1} = (a \cdot x_i + c) \bmod N \tag{3}$$

mod N stands for the residue modulo N. To obtain a random number in the (0,1) range, one has to divide the result by

$$y_i = x_i/N \tag{4}$$

We have employed the following values of the constants in the work: a= 1029, c = 221591, N = 1048576, $x_0 = 0$. These parameters were given in Reference 3. The overall period of this RNG was found to be 1,048,577.

The second RNG was taken from the BOSS molecular simulations program.[1g] It uses the same formula as in Equation 3, but there is an additional randomization of value of the parameter c with a secondary and a tertiary generators. This randomizing modification occurs after a number of steps modified with the secondary LCG, and by an amount defined by the tertiary generator. The period of this generator is 2,276,956.

The third generator we used is the Mersenne Twister generator by Matsumoto and Nishimura.[4] The complete description of the technique is given in Reference 4. Briefly, this algorithm generates a sequence of machine word vectors (with each element being either 0 or 1) $\mathbf{x_i}$ (with the dimensionality of w) according the Equation 5;

$$\mathbf{x_{k+n}} = \mathbf{x_{k+m}} \oplus (\mathbf{x_k^u} | \mathbf{x_{k+1}^j}) A, (k=0, 1, \ldots) \tag{5}$$

where n, m, and r are integer constants, A is an integer constant matrix with elements 0 and 1, $(\mathbf{x_k^u} | \mathbf{x_{k+1}^j})$ is the concatenation (a word vector obtained by concatenating the upper w-r bits of $\mathbf{x_k}$ and the lower r bits of $\mathbf{x_{k+1}}$, in this order), and $\oplus$ is bitwise addition modulo two. $0 \leq r \leq w-1$, $1 \leq m \leq n$. In the FORTRAN implementation of this random number generator which we used, the period of the pseudo-random sequence was equal to 884,585,892.

There was also a modification of the LCG used in this study. There is a known 3-D correlation between the numbers produced by this generator.[5] If one generates a sequence of the LCG pseudo-random numbers $\{x_i\}$ and then arranges them in triples as (x_1, x_2, x_3), (x_4, x_5, x_6), etc., then these triples can be displayed as points in 3-D Cartesian coordinates. It so happens that all these points belong to 2-D planes. Figure 2 shows this phenomenon for 100,000 points generated with the unmodified LCG. The 2-D randomization inside these

planes is still in place, as can be seen on Figure 3 which shows the same pseudo-random points from a different view angle. While it is not quite clear whether such a property would affect Monte Carlo simulations (the points do belong to these planes, but two adjacent triples do not necessarily belong to the same plane). However, it is obvious that there is a significant part of the space that is not covered by the points; therefore, there is a definite correlation between the numbers in this pseudo-random sequence.

The other two generators we tested are free from the above deficiency.

In order to alleviate the above problem, we have also studied a modified LCG, in which a shuffling was performed to remove the correlation. Whenever a random number was required in a Monte Carlo run, we used the following algorithm:

    a.   a random number was obtained with the LCG;

    b.   we determined the fourth decimal in this random number and ran the LCG 0 to 9 more times, according to this decimal. For example, if the initial random number was 0.978612, we performed 6 such runs;

    c.   after that, we generated another LCG random number, and used it in the Monte Carlo program.

Therefore, we shuffled the pseudo-random numbers generated by the LCG, and the correlation shown on Figure 2 was completely removed. However, the period of the generated sequence was reduced from 1,048,577 to 161,945.

Overall, the following random number generators were used in this work:

    a.   The Linear Congruential Generator with a= 1029, c = 221591, N = 1048576, x_0 = 0. We will further denote it as LCG. It was used in simulating pure liquid butane and methanol and in simulating of the solvated tridecaalanine peptide;

    b.   The Linear Congruential Generator with additional randomization, as used in the BOSS program. We will denote it as BOSS. It was used in simulating pure liquid butane and methanol and in simulating of the solvated tridecaalanine peptide;

    c.   The Mersenne Twister generator by Matsumoto and Nishimura. We will denote it as MTG. It was employed in simulating pure liquid butane and methanol;

    d.   The shuffled generator or SHUFF, as described in the paragraph immediately above. This generator was used to simulate pure liquid butane and methanol;

    e.   In case of the model labeled as ACC, the LCG (as in (a) above) was used to determine attempted Monte Carlo moves and the BOSS (as in (b) above) was employed in determining whether the move passes the Metropolis acceptance test. This generator was used to simulate pure liquid butane and methanol;

    f.   We also tested versions of the Linear Congruential Generator with different values of the parameters a, c, N, and the seed x_0 in simulating pure liquid butane. In this case, values of all these parameters are explicitly listed in the corresponding table.

It should also be noted that, whenever a modifier "rigid" is used, the butane or methanol molecules were kept rigid, with no internal motion allowed, throughout the simulations.

B. Monte Carlo Simulations of Pure Liquid Butane and Methanol

In each case, simulations were carried out with 216 molecules in a cubic cell with the periodic boundary conditions. NPT ensembles were used (constant pressure and temperature) with t = −0.5 °C for the pure butane (the normal boiling temperature) and 25

°C for the methanol runs. The pressure was set to 1 atm. The OPLS-AA force field was employed.1a Briefly, the interatomic interactions were evaluated with the following formula:

$$E_{nb} = \sum_{i<j} [q_i q_j e^2 / r_{ij} + 4\varepsilon_{ij}(\sigma_{ij}^{12}/r_{ij}^{12} - \sigma_{ij}^6/r_{ij}^6)]f_{ij}$$

(6)

The summation runs over all the pairs of atoms $i < j$ on molecules A and B or A and A for the intramolecular interactions. Moreover, in the latter case, the coefficient f_{ij} is equal to 0.0 for any i-j pairs connected by a valence bond (1–2 pairs) or a valence bond angle (1–3 pairs). $f_{ij} = 0.5$ for 1,4-interactions (atoms separated by exactly 3 bonds) and $f_{ij} = 1.0$ for all the other cases. This part of the force field was augmented with harmonic bond stretching and angle bending and Fourier torsion coefficients. The standard OPLS-AA parameter values were employed. Moreover, we simulated both liquids with completely flexible molecules and with rigid molecules, in the latter case all the intramolecular degrees of freedom were frozen. Intermolecular interactions were truncated at 11 Å. Ranges of molecular, atomic, and volume moves were adjusted to produce a ca. 30–60% acceptance ratio, although other options were also employed as described below. Volume moves were attempted once in every 600 steps. In each case, 1×10^6 Monet Carlo configurations of initial equilibration were followed by 4×10^6 configurations of data collection, both carried out in 2×10^5 configurations increments. All the simulations were carried out with the BOSS Version 4.6 program1g which was modified to use the nonstandard RNG – the LCG, shuffled LCG, and the Mersenne Twister generator by Matsumoto and Nishimura.

C. Simulation of Hydrated Tridecaalanine Peptide

The peptide simulations were carried out with the LCG and BOSS random number generators. The polypeptide chain was initially in the fully extended conformation. The peptide was solvated by the explicit water using the flexible four-site model available in the BOSS package.1g The number of water molecules was 1,796, which ensured a ca. 10 Å layer of water on each side of the tridecaalanine molecule at the beginning of the simulations. The isothermal-isobaric (NPT) ensemble was used in these Monte Carlo runs at 25 °C and 1 atm. Overall, 6×10^6 configurations we employed in each run. A 7.5 Å cutoff distance was applied for the solvent-solvent and solvent-solute interactions, with a 100 Å cutoff distance for the intra-solute ones. The LCG and BOSS random number generators were used. In order to ensure consistency of the results and to examine the effect of the initial structure, cross-simulations were also carried out. That is, after the initial 6×10^6 Monte Carlo configurations, the LCG system was simulated for another 5×10^6 configurations using the BOSS RNG, and vice versa.

III. Results and Discussion

A. Pure Liquid Butane

Table 1 displays results of simulating the pure butane liquid using the original BOSS, the LCG, and the Mersenne Twister (MTG) random number generators. The following data are reported: the average volume and the average total energy, the volume and total energy from the last 2×10^5 Monte Carlo configurations step, the overall acceptance ratio and the acceptance ratio for volume moves only. The energy and volume are reported for the whole system of 216 molecules.

Several conclusions can be made from the data in table 1. The BOSS and the MTG lead to very close values of the volumes (35343.367 Å^3 and 35411.525 Å^3, respectively, for the

flexible model) and to a reasonably close total energies (1558.825 kcal/mol and 1633.077 kcal/mol, again, for the flexible model). But using the Linear Congruential Generator clearly leads to an expansion of the liquid, with the average volume being about 5% greater on the 4×10^6 Monte Carlo configurations used in the data collection (37028.151 Å^3 for the flexible model). Given the increase of the total energy of the system (as simulated with the LCG) by about 1800 kcal/mol compared to the other two generators, we can conclude that most of the change happens in the intramolecular energy component. At the same time, even the data for the completely rigid molecules demonstrates an expansion if the LCG is employed, even though the energy and volume changes are less dramatic. Furthermore, the ranges for the attempted molecular and volume moves were the same for the three generators. While the overall acceptance ratio was similar regardless of the generator which was used, the acceptance of volume moves was higher in the case of the LCG. This can probably be explained as follows. Employing the Linear Congruential Generator leads to sampling of higher-energy regions of the configuration space. The extra "accumulated" energy creates non-equilibrium conditions, and the excess of the energy is released by expanding the whole system. We have run simulations in which the energy was printed at each step in order to determine if there were any sudden jumps or perhaps indications of a bug in the software. No such indications have been found. The energy grows smoothly and without any dramatic sudden changes.

It should also be noted that the BOSS and MTG runs produce the resulting volume in a closer agreement with the experimental numbers (34624.8 Å^3) than the LCG ones. Direct comparison of the total energies to the experimental values is not possible. However, the BOSS generator is known to produce the heat of vaporization for pure liquid butane at 5.44 kcal/mol, which is very close to the experimental number of 5.35 kcal/mol.[1a] Therefore, it is not unreasonable to imagine that the over-expanded liquid produced in the LCG run would have an underestimated enthalpy of vaporization.

Another interesting fact is that, even though the volumes of the pure liquid butane obtained with the original BOSS and the Mersenne Twister generators are virtually identical, the total energies for the flexible molecules differ by about 0.3 kcal/mol per one molecule. Since the intermolecular energies (the total energies in the rigid molecular case) are practically the same, the difference is most likely lying in the intramolecular sampling. Thus, even though these two generators produce essentially the same physical results, the energy still depends on the RNG employed, although in a much more subtle way than in the case of the LCG.

In order to further assess the sources of the above effects, we have performed calculations with two more versions on the Linear Congruential Generator. First, we tried the shuffled version of this generator, with the pseudo-random sequence produced as described in Section II A above (SHUF). Second, we used the LCG for the attempted molecular movements, but employed the standard BOSS RNG for the Metropolis acceptance test (ACC). Table 2 presents results of using these two generators. Apparently, the shuffling of the RNG pseudo-random sequence only makes the physical results worse. The energy grows by more than 400% for the flexible molecules (reaching 23363.805 kcal/mol) and by 43% for the rigid molecules (−897.262 kcal/mol). This leads to a rapid expansion of the system (which still continues at the end of the 5×10^6 configurations run) and to a very high acceptance ratio for the volume moves (93% for the flexible system). As was noted above, the period of such a shuffled generator (SHUFF) is only 161,946, much shorter than the original LCG period of 1,048,577. To check if the period length itself would be the reason for the problem, we performed a similar run with the standard BOSS RNG, which was shuffled in the same way (as described in Section II A). The period of the BOSS generator was thus reduced from 2,276,956 to 397,389. As a result, the average volume and energy were found to be 35124.737 Å^3 and 1525.151 kcal/mol, very close to the original BOSS

properties, with the volume moves acceptance rate of 45%. Therefore, we can conclude that (i) a short length of the period itself is not to be blamed for the unphysical results obtained with the LCG (although a shorter period can make the results worse) and (ii) the intramolecular energy is certainly the main contributing factor, but the expansion occurs even for the liquids composed of rigid molecules, for as long as the LCG is employed.

Examining the ACC results demonstrates that there is still a difference from the standard BOSS and MTG results, but the magnitude is smaller. For example, the volume of the flexible system is still more than 2% higher than the volume computed with the BOSS RNG, at 36575.790 $Å^3$. This leads us to a conclusion that performance of a random number generator is important in both the process of determining attempted molecular and atomic movements and the process of performing the Metropolis acceptance test.

The deviation of the volume from the experimental value is also smaller for the ACC simulation, although the result is still less accurate than the BOSS and MTG results.

There is another aspect of the Monte Carlo technique which we studied with respect to the influence of the random number generators involved. It is known that the recommended acceptance ratio is between ca. 30–60%.[6] The ratio can be adjusted by changing the ranges of attempted molecular and volume moves. Tables 1 and 2 list results obtained with the acceptance ratios close to the recommended (and achieved with the same attempted movement sizes regardless of the particular RNG). We have also attempted runs with larger magnitudes of the attempted intramolecular atomic displacements to test behavior of the system as simulated with the different random number generators in the non-perfect acceptance ratio region. The results are shown in Table 3.

One can make several interesting observations from the results in this table. First of all, the volume acceptance ratios for the BOSS (44%), LCG (52%), MTG (46%), and ACC (54%) remain largely unaffected by the ca. seven-fold change in the acceptance rates for the molecular motion (this is only true for the ACC model to a degree). At the same time, the shuffled generator (SHUFF) had the rate of rejecting the volume moves increased from 7% to 46%. This can be explained as follows. The SHUFF had the fastest energy growth in the case of the normal acceptance rate. This energy was mostly the intramolecular one, leading to the fastest observed volume expansion. As the volume of the system became very large (more than 5 times larger than with the original BOSS and the MTG generators), subsequent changes of volume were much less hindered by the intermolecular repulsion, and so virtually all the volume moves were accepted, although with a clear bias toward increasing the volume. Once the overall acceptance ratio was reduced, the intramolecular energy could no longer grow so fast as before, and thus the expansion was slower, with the smaller overall volume and a relatively high probability of rejecting the volume moves.

The second observation is that the results of applying the BOSS, LCG, and MTG generators seem to be rather close to those obtained within the normal acceptance ratio conditions. All the three resulting average volumes are also comparably close to the experimental number. There are minor changes – the overall energy for the BOSS and MTG cases is about 2% lower (which is to be expected when the acceptance ratio is lower). The volumes are very close to the original values. The Linear Congruential Generator produces a lower volume and less of an energy increase, which is also understandable for the same reason the SHUFF results are less different from the BOSS and MTG ones. Overall, we can conclude that the Monte Carlo runs with both the BOSS and MTG random number generators are fairly stable with respect to the molecular motion acceptance rates.

The results for the ACC (the Linear Congruential Generator with the BOSS generator used for the Metropolis acceptance test) are somewhat surprising. While the volume is not

changing much, the total energy in this case is much higher than for the overall acceptance rate of 45%. It is not clear to what this result should be attributed.

B. Pure Liquid Methanol

Having studied the behavior of the pure liquid butane as simulated with the various random number generators, we proceeded to a somewhat different system – pure liquid methanol. The liquid butane at its boiling temperature of −0.5 °C could be a volatile system, which would only need a small push for an expansion to start, while liquid methanol at 25 °C has a greater energy of intermolecular binding, and thus could be assumed to be more stable with respect to the choice of the random number generator employed in the Monte Carlo simulations.

Presented in Table 4 are results of simulating pure liquid methanol with the BOSS, LCG, and MTG. We can see that the expansion produced by employing the Linear Congruential Generator is actually greater in this case than for the butane calculations. The average volume of the system as computed with the LCG is 18636.744 Å^3 vs. the ca. 15000 Å^3 as obtained with the BOSS and MTG, an increase of 24%. Moreover, it is obvious that the final average volume would be even greater, as the volume at the last 2×10^5 configurations is even higher (20307.890 Å^3). At the same time, the BOSS and MTG runs produce similar results once again, although the energy obtained with the Mersenne Twister generator is slightly higher (−104.038 for the flexible model, compared to −129.962 kcal/mol for BOSS).

The BOSS and MTG runs produce the resulting volume in a closer agreement with the experimental value of 14623.2 Å^3 than the LCG ones. Direct comparison of the total energies to the experimental values is again not possible. But the BOSS generator is known to produce the heat of vaporization for pure liquid methanol of 8.95 kcal/mol, which is exactly the same (within the precision) as the experimental value.[1a] Therefore, using the LCG generator is likely to reduce the accuracy of the methanol heat of vaporization.

Table 5 lists data obtained for the pure liquid methanol with the shuffled LCG (SHUFF) and the ACC (Linear Congruential Generator for the attempted moves, BOSS generator for the Metropolis acceptance test). The effect of using these generators is similar to that in the butane case, although it is even more profound. For example, the average volume with the SHUFF, which is 101125.270 Å^3 for the flexible model, increased by about a factor of 6.7 compared to the results obtained with the standard BOSS and MTG generators. Moreover, even freezing the intramolecular degrees of freedom does not help very significantly, with the total volume averaged over the last 4×10^6 Monte Carlo configurations being almost 3 times greater than the correct value (43484.392 Å^3). The ACC version of the pseudo-random sequence produces an effect somewhere between the LCG and the original BOSS generators, very similarly to the results in the pure liquid butane case described above. The average volume for the flexible model is 15578.559 Å^3, and the one for the rigid molecule simulation is a bit higher at 15688.341 Å^3.

Just like in the case of pure liquid butane, the ACC model seems to produce a total volume which is somewhat closer to the experimental value than the original LCG.

We have also studied effects of using larger attempted moves and thus a lower overall acceptance ratio, just like in the case of liquid butane. The results are given in Table 6 and the general trends are similar to those observed in the butane case. For example, using both BOSS and MTG generators leads to results which are very similar to those obtained in the normal acceptance range, although the energy tends to be slightly lower, while the average volume values for these models are the closest to the experimental one (14699.013 Å^3 and

14801.231 Å^3 for the BOSS and MTG, respectively, compared to the experimental value of 14623.2 Å^3). We can conclude that these two generators are stable with respect to changes in the acceptance ratio. At the same time, using not only the ACC but also the SHUF generator leads to an increase of the total energy as compared to the results obtained in the normal acceptance ratio region.

Overall, we can summarize the data presented in this subsection by stating that (i) the methanol results are generally similar to those obtained for the pure liquid butane and (ii) the magnitudes of the trends observed here are greater than those in the butane case. This is likely an indication that the methanol system is a better testing case for any new random number generator to be employed in Monte Carlo simulations, since deviations from the well-tested BOSS random number generator and the apparently well-behaved Mersenne Twister generator results are greater in this case. Therefore, if the new generator is likely to cause overexpansion (or overcontraction) of the volume and over- or underestimation of the total energy, this trend should be more pronounced for the pure liquid methanol. Therefore, they are easier to discover for this system, and using it as the test case will be more efficient.

The following general comment should be made regarding the fact that the energy of the systems is sometimes significantly higher when simulated with LCG as compared to the BOSS and MTG generators, even though the volume is not necessarily following this trend to the same degree. We believe that the key to this issue can be found, for example, in the data presented in Table 1. The energy for the rigid LCG simulation is −995.173 kcal/mol, only ca. 25 kcal/mol or 2.5% different from the BOSS and MTG versions of this property. In the case of a rigid simulation, practically all the energy comes from the intermolecular interactions. And a relatively modest increase in volume corresponds to a modest increase in energy. However, when the molecules are completely flexible, the overall energy with LCG is almost 2000 kcal/mol higher than that in the BOSS or MTG model. This means that most of this increase is in the internal energy of the individual molecules, which does not necessarily manifests significantly in the intermolecular part of the total energy, and thus does not affect the volume in a proportionally significant fashion.

C. Tridecaalanine Peptide Simulations

We have simulated this solvated peptide in order to assess possible effects of the random number generator choice upon simulations of biologically relevant systems. Tridecaalanine is a small peptide which still exhibits basic features of alpha-helix in aqueous solution. Shown on Figure 4 are volumes of the BOSS– and LCG–simulated hydrated extended tridecaalanine peptide systems as a function of the number of Monte Carlo steps.

Both simulations start from exactly the same configuration. It is clear from the graph on Figure 4 that they converge for very different volumes. The BOSS generator leads to ca. 60000 Å^3 volume while the LCG one yields about 110000 Å^3, almost two-fold greater. At the same time, the total energy of the solvated peptide system levels off at about −5000 kcal/mol if the Linear Congruential Generator is used, and the BOSS random number generator produces the total energy of at least −13000 kcal/mol. This behavior of the total energy is shown on Figure 5.

The pattern of the LCG giving an overestimated volume, as compared to the results of simulating with the BOSS random number generator, is consistent with the trends in the neat liquid simulations described in the previous subsections. Moreover, the difference in volume comes from inflating the volume of the solvent and not from a significant change in the structure of the solute. This is natural given that the effects we are observing seem to depend on some very fine balance of parameters, and it would take a more profound effect to

drastically change the geometry of a polypeptide to a degree that could account for such a significant volume expansion.

Shown in Figures 6 and 7 are snapshots of the solution geometries. It can be clearly seen that the LCG simulation leads to emergence of cavities in the solvent which are responsible for the significantly greater size of the system.

Let us now consider what happens if we switch the random number generators after the initial 6×10^6 Monte Carlo configurations. That is, as described in the previous section, we continue the BOSS RNG – equilibrated simulations with the LCG random number generator and vice versa. The effect of this RNG switching on the volume and energy of the system is shown on Figures 8 and 9, respectively.

These results are rather interesting. The BOSS random number generator permits for both the volume and the energy to return to their true values (it is very reasonable to assume that the BOSS results are the "real" Monte Carlo ones given the overwhelmingly successful history of applying this software). Moreover, the LCG generator does not significantly change the volume, and the energy, although rising to ca. −10000 kcal/mol, is still two times lower than the purely LCG-generated value of ca. −5000 kcal/mol. Apparently, once the system is equilibrated with the BOSS RNG, it gains a certain stability which is not so easily upset by applying the LCG generator. Therefore, the effect of choosing a specific random number generator, while physically significant, appears to be based on a very delicate balance of influences.

In order to test the significance of the presence of the peptide, we have also run simulations of 216 molecules of pure water. After 1×10^6 Monte Carlo configurations of averaging follows by 5×10^6 configurations of averaging, the LCG volume was 8114.6 $Å^3$, or only about 0.6 % greater than the BOSS result of 8064.9 $Å^3$. The experimental value of the volume is 8010 $Å^3$.1a The LCG and BOSS energies were −2186.8 kcal/mol and −2207.8 kcal/mol, respectively. Therefore, the dramatic effect of the choice of a random number generator on the properties of the solutions is indeed caused by the presence of the peptide.

D. Testing Other Values of Parameters and Conversion

In order to test the effect of using different seed (x_0 in Table 7) values of parameters a, a, and N in the Linear Congruential Generator, we carried out calculations for the pure liquid butane systems using a variety of the values. First of all, different sets of parameters a, c, N, and the seed x_0 were tested. In these calculation, 1×10^6 Monet Carlo configurations of initial equilibration were followed by 4×10^6 configurations of data collection, both carried out in 2×10^5 configurations increments. The results are shown in Table 7. It can be seen from the first four rows of data in that table that the choice of the seed (x_0) for the LCG does not significantly affect the outcome of the tests. At the same time, using different sets of the a, c, and N parameters can have a very noticeable effect. The average volume varied from 37028.15 $Å^3$ for the LCG from the Allen and Tildesley book3 down to 35225.15 $Å^3$ for LCG from Reference 8, and with a very large differences in energy. Given these results, we can conclude that the the Linear Congruential Generator is in principle capable of performing well in Monte Carlo simulation, but only with certain values of the a, c, and N parameter, and these values cannot be predicted from statistical tests only, but rather require actual Monte Carlo simulations to be carried out.

Table 8 shows the average volumes obtained with the Linear Congruential generator with the parameters from Reference 3 after different numbers of Monte Carlo configurations. Averaging was always done for the last 4×10^6 configurations. It can be easily seen that both the average volume and the average energy do not experience any long-range drift and

are essentially converged after the very first four million Monte Carlo configurations of averaging, with the average volume ca. 36960 Å^3. It evolves from 37028.151 Å^3 for 5×10^6 configurations to 36765.859 Å^3 for 21×10^6 configurations, and the change is not uniform but rather oscillatory, which confirms that the simulations have converged.

E. Testing Uniformity and Average Values of the Pseudo-Random Numbers

We have shown above that the poor performance of the Linear Congruential Generator cannot be explained by its relatively short period alone. In order to further explore the issue, we have studied the uniformity of the BOSS, LCG and MTG random number generators using the technique described in Section I above. We used different lengths of the sequences, and the resulting maximum deviations from the uniform line were averaged over 100 consecutive pseudo-random runs in each case. Shown in Table 9 are these averaged maximum deviations for the different generators calculated for different sequence lengths and with the different generators.

It is clear that the numbers for all the sequence lengths presented (from 10 to 10000) are practically identical regardless of what random number generator is used. Therefore, we can conclude that the uniformity of the sequence is not the reason for the poor physical results obtained in Monte Carlo simulations with the Linear Congruential Generator.

Since the uniformity test does not indicate whether deviations are equally probable to be in the positive and negative directions, we have also calculated the average values of the pseudo-random numbers produced by the three generators. These average values are displayed on Figure 10 as a function of $\log_{10}$ of the number of steps in the sequence.

It can be seen that in this case there is actually a difference in the BOSS and MTG generators on one hand, and the LCG on the other hand. In the range between ca. 1,000 and 40,000 steps, the LCG has a higher deviation from 0.5 than the two "good" generators. Moreover, the deviation is negative. However, the magnitude of the difference is quite small, less than 0.01.

As described in the previous subsection, there are values of the LCG parameters which can remedy the deficiency seen in the generator behavior. We have studied the average properties of the generators from References 6⁻8. The results for the Numerical Recipes parameters6 are given on Figure 10 along with the BOSS, LCG and MTG results. The outcome of using the parameters from References 7 and 8 were similar, with the exception of the averages being closer to 0.5 after ca. $10^{4.2}$.

It is clear from the picture, that the "good" LCG differs from the "bad" LCG by having a smaller deviation from the 0.5 average in the range of ca. 1,000–30,000 steps. This is very similar to the difference in the behavior between the "bad" LCG and the BOSS and MTG generators. This advantage of the Numeric Recipes generator is apparently not offset even by its being less close to 0.5 for longer sequences. Therefore, there is a possibility that the problem with the LCG generator is caused by its average values not being close enough to the 0.5 value in the important range of 10,000–30,000 randum number generator calls.

The above statement can be illustrated by data on Figure 11, which is built similarly to Figure 1. 1000 sequences of pseudo-random numbers were generated with the BOSS, MTG and LCG techniques. Then the numbers were sorted and each was plotted versus its position in the sorted sequence. It can be seen that the LCG sequence does demonstrate a slight trend to have the numbers of lower value as compared to both BOSS and MTG ones.

IV. Conclusions

We have carried out Monte Carlo simulations of pure liquid butane and methanol using three different random number generators – the standard RNG in the BOSS program, the Mersenne Twister generator (MTG) by Matsumoto and Nishimura, and the standard Linear Congruential Generator, LCG, (including some variations). We have also simulated aqueous solution of tridecaalanine peptide with the BOSS and LCG random number generators. While employing the BOSS and MTG generators leads to very similar physical results, the LCG with the values of parameters from Reference 3 produces and expansion of the liquids, which is more pronounced in the case of methanol and reaches ca. 24%. The discrepancy is made even worse if a simple shuffling of the LCG pseudo-random sequence is carried out. Moreover, the solvated peptide simulations with the LCG code resulted in a volume of about twice as big as produced by the BOSS random number generator in a significantly less negative energy of the system. At the same time, cross-switching the generators after the initial 6 million of Monte Carlo configurations lead to convergence to more similar results.

All the generators employed in this study have almost identical uniformity of the generated sequences, although the average values tend to deviate more from 0.5 for the original LCG in the 1,000–30,000 steps range. It has also been shown that certain changes in the LCG parameter values are capable of remedying the problem. Thus, we can conclude that, on one hand, the choice of a random number generator for Monte Carlo simulations can have a very strong effect on the physical results of the calculations and, on the other hand, simple statistical tests are not sufficient to identify all the potential problems. It can therefore be recommended that a random number generator should be tested by simulating a well-studied system, such as the pure liquid methanol at 25 °C, and with a well-established force field, before being employed in routine Monte Carlo simulations.

Acknowledgments

References

1. For representative publications see: (a) Jorgensen WL, Maxwell DS, Tirado-Rives J. J Am Chem Soc. 1996; 118:11225–11236. (b) Kaminski G, Jorgensen WL. J Phys Chem. 1996; 100:18010–18013. (c) Tominaga Y, Jorgensen WL. J Med Chem. 2004; 47:2534–2549. [PubMed: 15115396] (d) Zhou J, Tsao HK, Sheng YL, Jiang SY. J Chem Phys. 2004; 121:1050–1057. [PubMed: 15260639] (e) Kamberaj H, Van der Vaart A. J Chem Phys. 2007; 127:234102. [PubMed: 18154370] (f) Kim YC, Hummer G. J Mol Biology. 2008; 375:1416–1433.(g) Jorgensen, WL. BOSS, Version 4.6. Yale University; New Haven, CT: 2004.

2. Dwyer, Jerry; Williams, KB. Testing Random Number Generators. [Accessed March 27, 2008]. http://www.ddj.com/cpp/184403185

3. Allen, MP.; Tildesley, DJ. Computer Simulations of Liquids. Oxford University Press Inc; New York, NY: 1989.

4. Matsumoto M, Nishimura T. ACM Transactions on Modeling and Computer Simulation. 1998; 8:3–30.

5. Matsumoto M, Saito M, Haramoto H, Nishimura T. J Universal Computer Sci. 2006; 12:672–690.

6. Press, WH.; Flannery, BP.; Teukolsky, SA.; Vetterling, WT. Numerical Recipes in FORTRAN 77: The Art of Scientific Computing. 2. Cambridge University Press; 1992.

7. Microsoft Visual/Quick C/C++. [Accessed September 23, 2008]. http://msdn.microsoft.com/en-us/visualc/default.aspx

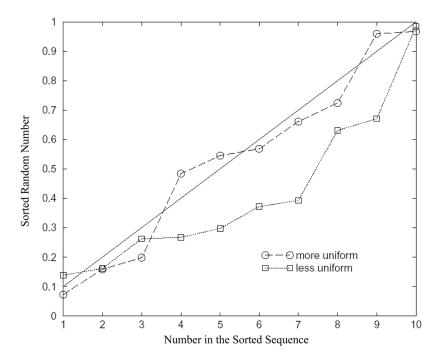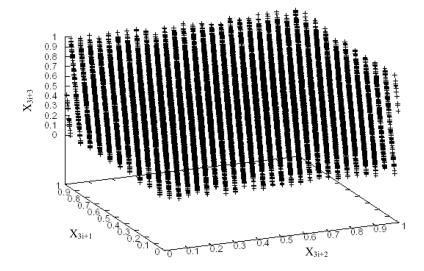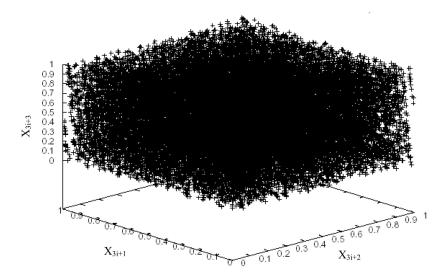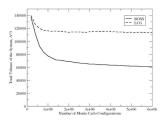8. Park SK, Miller KW. Comm Of the ACM. 1988; 31:1192–1201.

Figure 1.
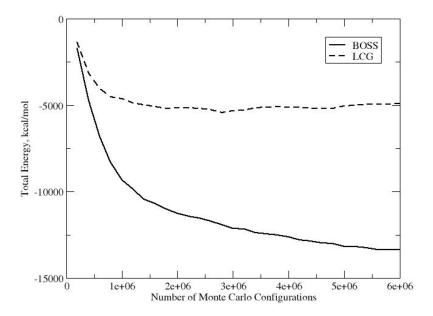A straight line y = x and two sorted sequences of pseudo-random numbers, a more uniform and a less uniform one.

Figure 2.
Triples of consequent pseudo-random numbers produced with the LCG represented as points in 3-D Cartesian coordinates. It can be seen that the points are arranged in 2-D planes.
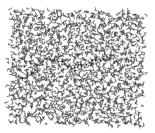
Figure 3.
Triples of consequent pseudo-random numbers produced with the LCG represented as points in 3-D Cartesian coordinates. At this view angle the points appear to be distributed randomly.
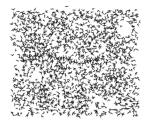
Figure 4.
Total volume of the hydrated extended tridecaalanine as a function of number of Monte
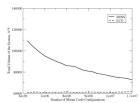Carlo configurations.

Figure 5.
Total energy of the hydrated extended tridecaalanine as a function of number of Monte Carlo configurations.
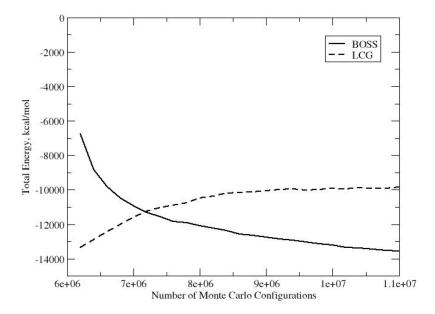
Figure 6.
A snapshot of the tridecaalanine system simulated with the BOSS random number generator.

Figure 7.
A snapshot of the tridecaalanine system simulated with the LCG random number generator.

Figure 8.
Total volume of the hydrated extended tridecaalanine as a function of number of Monte Carlo configurations after switching the random number generators. The RNG named here are the ones used in continuation of the calculations.

Figure 9.
Total energy of the hydrated extended tridecaalanine as a function of number of Monte Carlo configurations after switching the random number generators. The RNG named here are the ones used in continuation of the calculations.
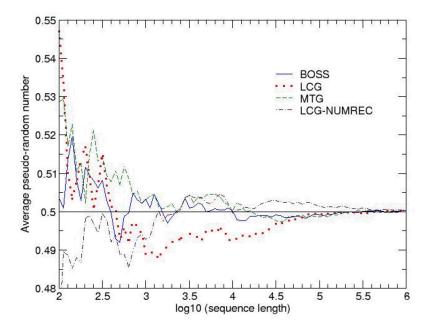
Figure 10.
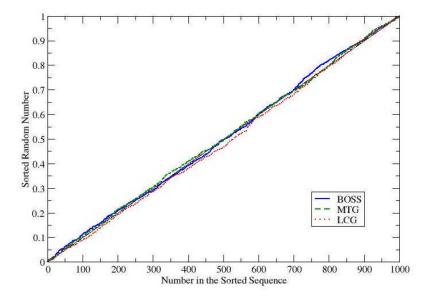Averaged pseudo-random numbers as functions of $\log_{10}$ of the sequence length.

Figure 11.
A straight line y = x and sorted sequences of pseudo-random numbers.

Table 1

Results of Simulating Pure Liquid Butane with the BOSS, LCG, and MTG Random Number Generators. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
BOSS	34366.747	35343.367	1521.245	1558.825	34%	43%
BOSS rigid[a]	36071.634	35811.822	−1016.629	−1021.280	50%	48%
LCG	37871.455	37028.151	3529.672	3465.947	34%	52%
LCG rigid[a]	37851.728	36840.426	−967.128	−995.173	50%	55%
MTG	35071.625	35411.525	1604.410	1633.077	34%	46%
MTG rigid[a]	36045.273	35821.250	−1012.936	−1022.023	50%	46%
Experiment[1][a]		34624.8				

[a]No intramolecular motion is allowed.

Table 2

Results of Simulating Pure Liquid Butane with the SHUF (shuffled LCG) and ACC (LCG combined with BOSS) Random Number Generators. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
SHUF	182962.703	139311.149	26557.337	23363.805	39%	93%
SHUF rigid[a]	49306.569	41949.631	−808.672	−897.262	53%	59%
ACC	36563.171	36575.790	3165.301	3144.410	34%	45%
ACC rigid[a]	36001.988	36163.909	−1013.190	−1009.296	50%	50%
Experiment[1a]		34624.8				

[a]No intramolecular motion is allowed.

Table 3

Results of Simulating Pure Liquid Butane with the Ranges of Intramolecular Attempted Moves Adjusted to Decrease the Overall Acceptance Ratios. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
BOSS	34998.032	35501.188	1538.104	1527.953	5.6%	44%
LCG	36066.137	35862.969	3506.502	3080.883	5.5%	52%
MTG	35360.646	35204.808	1670.756	1616.600	5.5%	46%
SHUF	36724.393	36469.683	10790.760	7889.214	3.4%	46%
ACC	37130.026	36908.502	8985.923	6888.825	5.0%	54%
Experiment[1a]		34624.8				

Table 4

Results of Simulating Pure Liquid Methanol with the BOSS, LCG, and MTG Random Number Generators. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
BOSS	14798.584	14851.023	−171.377	−129.962	44%	47%
BOSS rigid[a]	14798.903	14955.402	−1737.385	−1723.882	66%	45%
LCG	20307.890	18636.744	237.433	200.756	46%	59%
LCG rigid[a]	16178.770	16409.558	−1636.942	−1626.920	67%	50%
MTG	14576.476	15115.916	−154.184	−104.038	44%	45%
MTG rigid[a]	14999.360	15017.511	−1705.676	−1695.512	66%	50%
Experiment1[a]		14623.2				

[a]No intramolecular motion is allowed.

Table 5

Results of Simulating Pure Liquid Methanol with the SHUF (shuffled LCG) and ACC (LCG combined with BOSS) Random Number Generators. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
SHUF	129422.653	101125.270	2505.412	2396.234	51%	89%
SHUF rigid[a]	52694.685	43484.392	3375.077	3350.707	15%	84%
ACC	15578.559	16401.748	−3.177	55.145	45%	44%
ACC rigid[a]	15688.341	15365.562	−1635.879	−1670.544	66%	47%
Experiment[1a]		14623.2				

[a]No intramolecular motion is allowed.

Table 6

Results of Simulating Pure Liquid Methanol with the Ranges of Intramolecular Attempted Moves Adjusted to Decrease the Overall Acceptance Ratios. Volumes are in Å^3, energies in kcal/mol.

Method	V(last)	V(average)	E(last)	E(average)	Acceptance	Acc. for V
BOSS	14330.210	14699.013	−209.336	−166.649	12%	37%
LCG	16193.728	16132.413	304.082	340.087	12%	46%
MTG	14441.051	14801.231	−139.102	−109.429	12%	45%
SHUF	52694.685	43484.392	3375.077	3350.707	15%	84%
ACC	15123.067	15440.551	305.848	324.816	12%	48%
Experiment[1a]	14623.2					

Table 7

Results of Simulating Pure Liquid Butane with the LCG Using Different Parameter Values. Volumes are in Å^3, energies in kcal/mol.

a	c	N	X$_0$	V(last)	V(average)	E(last)	E(average)
1029	221591	1048576	0	37871.46	37028.15	3529.67	3465.95
1029	221591	1048576	1	36213.46	36856.84	3650.55	3498.69
1029	221591	1048576	100	36059.55	36427.05	3437.47	3516.19
1029	221591	1048576	10000	36006.11	36909.08	3503.84	3519.71
1664525[a]	1013904223	4294967296	0	35628.90	35648.10	1560.92	1550.13
214013[b]	2531011	4294967296	0	35304.82	35515.08	1546.77	1538.79
16807[c]	0	2147483647	1	34804.56	35225.15	1553.20	1534.07
Experiment1[a]					34624.8		

[a]Reference 6.

[b]Reference 7.

[c]Reference 8.

Table 8

Results of Simulating Pure Liquid Butane with the LCG after different numbers of configurations. Volumes are in $Å^3$, energies in kcal/mol.

Number of Monte Carlo Configurations	V(average)	E(average)
5×10^6	37028.151	3465.947
9×10^6	36934.985	3527.905
13×10^6	36794.495	3553.196
17×10^6	37270.543	3546.526
21×10^6	36765.859	3569.537
Experiment[1a]	34624.8	

Table 9

Averaged maximum deviations from the uniformity for pseudo-random sequences of different lengths. Volumes are in Å^3, energies in kcal/mol.

Method/Length	10	100	1000	10000
BOSS	0.2084	0.0747	0.0266	0.0089
LCG	0.2078	0.0768	0.0275	0.0086
MTG	0.2067	0.0769	0.0273	0.0084