VOLUME 120, NUMBER 29
JULY 29, 1998


JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

# Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid

**Carlos Simmerling,[†] Jennifer L. Miller,[‡] and Peter A. Kollman*[,†]**

*Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94148-0446, and CombiChem, Inc., 1804 Embarcadero Road, Suite 201, Palo Alto, California 94303*

*Received August 4, 1997*

**Abstract:** One of the largest challenges facing structural biologists is to predict the structure of complex molecules in aqueous solution. In this article we show for the first time that Locally Enhanced Sampling can be used in conjunction with the recently implemented Particle-Mesh-Ewald method to address this challenge. We present results from studies of a small RNA tetraloop system (5′-r[GGACUUCGGUCC]) where this combined protocol has allowed our simulations to rapidly find the experimental structure starting from an alternate location in phase space.

## Introduction

Structural biologists are concerned with the structure of biomolecules such as proteins and nucleic acids. Many structural studies are undertaken using the experimental techniques of NMR and/or X-ray crystallography. Under optimal conditions these experimental techniques return data which, after refinement, give a representation of the average structure of the molecule. In addition, these experiments may provide some insight into a molecule's motion in solution. Structure refinement of such data relies on the use of theoretical methods such as molecular mechanics and molecular dynamics (MD). None of these approaches are foolproof, and, in the case of nucleic acids, both experimental and theoretical methods face serious limitations,[1] making structure determination of this important class of molecules a particularly challenging problem.

For example, many nucleic acid sequences are difficult to crystallize, and those that do are subject to packing effects that may or may not represent their solution-phase structures. NMR-derived structures are increasingly available, but the NOEs observed during the experiment are limited in range to less than 5 Å. This limitation hinders efforts to study nonhelical nucleic acid structures in which regions that contain loops and bulges (regions important in many RNA-protein interactions) are frequently underdetermined by the experimental data.[2] MD, in addition to its use in refining X-ray and NMR structures, has also been used independently to investigate the structure and dynamics of nucleic acid structures in solution. However, simulations which do not begin near the experimentally determined structure, or where the experimental structure is ill-determined, have been limited both by the theoretical description of the system and by the amount of sampling achievable during a simulation.

Stable MD simulations of highly charged systems such as

* To whom correspondence should be addressed.
[†] University of California, San Francisco.
[‡] CombiChem, Inc.

(1) Liu, H.; Spielmann, H.; Ulyanov, N.; Wemmer, D.; James, T. *J. Biomolecular NMR* **1995**, *6*, 390−402.

(2) Battiste, J. L.; Mao, H. Y.; Rao, N. S.; Tan, R. Y.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. *Science* **1996**, *273*, 1547−1551.

S0002-7863(97)02702-9 CCC: $15.00 
Published on Web 07/11/1998

nucleic acids with explicit representation of solvent and counterions have only recently become feasible with the development of rapid but accurate methods for the treatment of long-range electrostatic interactions, such as the Particle-Mesh-Ewald[3] (PME) method implemented in AMBER.[4] With standard cutoff methods, unrestrained DNA and RNA simulations rapidly fall apart within a few hundred picoseconds.[5] However, even when an accurate force field allows for the stable simulation of native conformations, locating this conformation from alternate positions in phase space can be far from trivial.

Among the various techniques to increase sampling during a simulation, Locally Enhanced Sampling (LES)[6] stands out as a promising strategy. LES and related methods have been used in the past for a variety of problems.[6−16] This mean-field technique focuses the increased conformational sampling on a region of the structure. The "enhanced" sampling is accomplished through the use of multiple, but discrete, copies of the region of interest.

In the present study, we apply LES to a small RNA hairpin loop (5′-r[GGACUUCGGUCC]) which includes a four base pair stem and a four base tetraloop (Figure 1). The UUCG tetraloop provides an excellent system to test this combined LES/PME method for a number of reasons: there are two closely related NMR structures, the first one incorrect[17] (I) and the subsequent one correct[18] (C) and previous simulations of the RNA sequence have shown no interconversion between the two conformers in 2.5 ns of 300K MD.[19] Conversion from conformation I to C using MD was only observed when the 2'OHs were removed from the loop residues[19] and the simulations were carried out for the resulting RNA/DNA chimeric system. This chemical modification is not desirable in the general case of structure refinement, particularly in RNA systems where the ribose hydroxyls often participate in structure determining roles. Here we have used LES/PME on the native RNA sequence and, without employing any restraints, observed the conversion from the incorrect to the correct structure.

## Methods

The details of the LES approach have been described in detail in the past.[6,12,20,21] The method takes advantage of the observation that we are typically more interested in the conformational sampling of a subset of atoms in a simulated system. This is especially true for biomolecules in explicit solvents. Whereas most standard simulation methods expend a similar computational effort on all degrees of freedom, LES gives us the opportunity to focus additional resources on the interesting portion without significantly increasing the overall computational cost. This is accomplished by dividing the system into regions and replacing one or more regions of particular interest with multiple copies. The copies, which do not interact during the simulation, are free to move apart and explore different regions of phase space, thereby increasing the statistical sampling. It has also been shown[12] that the barriers to conformational transitions are reduced when using LES, resulting in a smoother potential energy surface which permits more frequent conformational changes. However, it is also important to note that that there is a direct and simple correspondence between the global potential energy minima of the standard and LES systems: the global minimum of the LES system occurs when all copies occupy the conformation corresponding to the global minimum of the original system.[12] One of the key advantages of LES over some other methods to enhance sampling is that it can be combined with molecular dynamics and an explicit representation of solvent.[10]

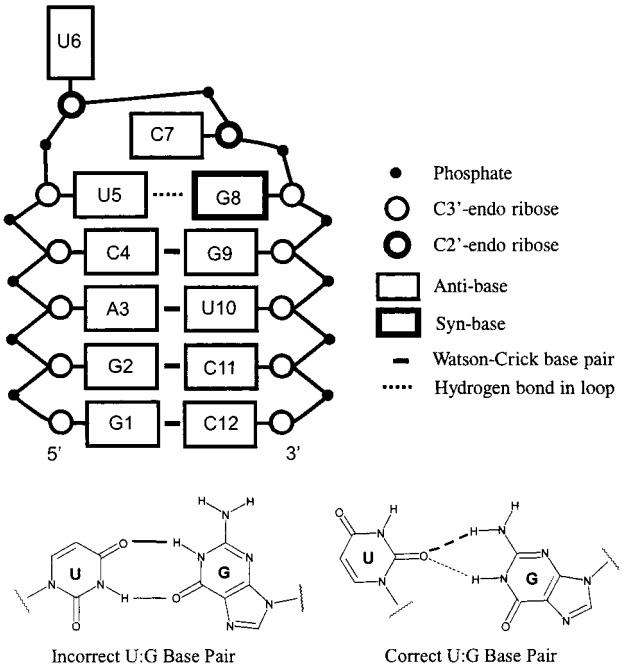


**Figure 1.** A schematic diagram of the topology of the RNA tetraloop system being studied. The hydrogen bond patterns for the U5:G8 base pair in the incorrect and correct NMR structures are shown in the lower figure. Solid lines are used for the incorrect hydrogen bonds, and dashed lines are used for the correct ones.

**Table 1.** Summary of the Results Obtained from the Various LES Simulations[a]

| LES region | no. of LES copies | total no. of particles | starting structure | time of I → C (ps) | final RMSD to C |
|---|---|---|---|---|---|
| UUCG[b] | 5 | 6874 | I | n/a | 1.8 |
| UUCG | 5 | 7358 | I | 175 | 0.9 |
| UUCG | 5 | 7358 | I | 200 | 0.7 |
| UUCG | 5 | 7358 | C | n/a | 0.9 |
| UUCG | 2 | 6983 | I | 600 | 1.1 |
| CUUCGG | 5 | 7618 | I | 100 | 1.2 |

[a] The non-LES system had 6858 atoms. [b] Only 2'HO atoms were copied.

(3) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(4) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Comm.* **1995**, *91*, 1−41.
(5) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194.
(6) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(7) Czerminski, R.; Elber, R. *Proteins* **1991**, *10*, 70−80.
(8) Li, H. Y.; Elber, R.; Straub, J. E. *J. Biol. Chem.* **1993**, *268*, 17908−17916.
(9) Verkhivker, G.; Elber, R.; Gibson, Q. H. *J. Am. Chem. Soc.* **1992**, *114*, 7866−7878.
(10) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534−2547.
(11) Simmerling, C. L.; Elber, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3190−3193.
(12) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277−9287.
(13) Zheng, Q.; Kyle, D. *J. Mol. Biol.* **1994**, *19*, 324−329.
(14) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *Biopolymers* **1996**, *39*, 103−114.
(15) Miranker, A.; Karplus, M. *Proteins* **1991**, *11*, 29−34.
(16) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisis, C. *J. Mol. Biol.* **1993**, *234*, 515−521.
(17) Varani, G.; Cheong, G.; Tinoco, I. J. *Biochemistry* **1991**, *30*, 3280−3289.
(18) Allain, F. H. T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.
(19) Miller, J. L.; Kollman, P. *J. Mol. Biol.* **1997**, *270*, 436−450.
(20) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380−3388.
(21) Straub, J. E.; Karplus, M. *J. Chem. Phys.* **1991**, *94*, 6737−6739.

In the present study we combined the LES approach with the PME method for calculation of long-range electrostatic interactions. This required changes to LES and PME as implemented in AMBER 5.0. The most important of these changes involved the intracopy nonbonded interactions, which will be described below.

The potential energy function for the LES system is constructed such that, when all copies occupy the same positions, the potential energy is identical to the corresponding single-copy (non-LES) system. As a result of this feature and fact that the copies have no direct interactions, the global minimum of this potential energy is unchanged from the original system.[12] For the covalent terms in the potential function (bond, angle, dihedral), this is achieved in the AMBER implementation by simply dividing the appropriate force constants by the number of times that the term was duplicated. For example, a bond that was duplicated $N$ times when LES is applied would have a force constant that is the original value divided by $N$. The sum of the bond energies for these copies is equivalent to the non-LES system when all of the bonds have the same length in both systems. When multiple LES regions are present, each copy in a particular region does not interact with other copies of the same region but interacts with all of the copies in the other regions. At the interface between different regions that are covalently linked, there are $N*M$ copies of these covalent terms for $N$ copies of region 1 and $M$ copies of region 2. These force constants are therefore scaled by $1/(N*M)$. All of this scaling is performed during the creation of the LES system and does not require recalculation during the simulation.

The nonbonded interactions are less simple. Once again, copies inside a particular region do not interact, while different regions involve an average interaction between all pairs of copies. The partial charges and Lennard-Jones well depth parameters for each atom are scaled by $1/N$ for $N$ copies. For interactions between different regions, this again results in a net scaling of the pairwise interactions by $1/(N*M)$. Since there are $N*M$ interactions between all of the pairs of copies, this provides the correct average energy. However, when atoms that are inside a particular region interact, this sum is incorrect and underestimates the magnitude of the interaction. The interactions are scaled by $1/N^2$, but since the copies do not interact there are only $N$ of these pairs, one for each copy (in contrast to the $N*M$ pairwise interactions for atoms in different LES regions). The correct scaling factor for the individual charges in this case would be $1/N^{1/2}$. If a simple Coulomb term is used for the electrostatic interactions, this means that LES atoms do not have a unique partial charge, but the value of this charge depends on whether the other atom is in the same LES region. Since the partial charges and well depths are precalculated for computational efficiency, intracopy nonbonded interactions need to be increased by a factor of $N$ during the simulation. In the non-PME LES implementation in AMBER, this scaling factor was directly applied to all intracopy nonbonded interactions.

When Ewald summation techniques are used, however, this scaling factor cannot be easily applied. The direct space calculation is similar to that performed when using a nonbonded cutoff. For the reciprocal space sum a single charge grid is constructed and used for the electrostatic potential therefore, each atom needs a unique partial charge value for contribution to this grid. The net electrostatic interactions for a particular pair of atoms involves contributions from both the direct and reciprocal space components, and this reciprocal space component will be incorrect for intracopy pairs. When the LES and PME methods were combined in AMBER, we added an explicit correction for these pairs. At the beginning of the simulation a list is constructed of all intracopy atom pairs, and the scaling factor of $N$ is used for their Lennard-Jones interactions. After each evaluation of the Ewald sum, the energies and forces from each of these intracopy pairwise interactions are reevaluated using a normal Coulomb calculation, scaled by the appropriate factor $(1/N - 1/N^2)$ and added to the energies and forces that were determined by the Ewald summation.

In addition, while each atom inside a region does not interact with other copies of the same region, these interactions are included for periodic images of the system. For example, atoms in copy 1 of a region do not interact with any atoms in different copies of the same region in the same molecule, but they do interact with all copies of this region for the periodic images of the system using the original $1/N^2$ scaling. This behavior is desirable and similar to that calculated for non-LES systems where atom pairs that do not have nonbonded interactions in

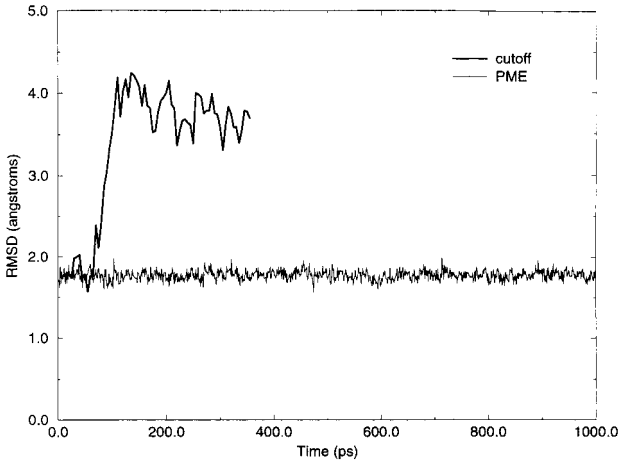


**Figure 2.** The loop RMSD (compared to the correct structure) as a function of time for simulations starting from the incorrect conformation. Five LES copies were employed for only the ribose hydroxyl hydrogen atoms in the loop. The thick line demonstrates that the loop is unstable when a cutoff is applied to nonbonded interactions. The thin line shows the data from a simulation using the PME method. Although the loop is stable, no improvement compared to the correct structure is observed during 1 ns.

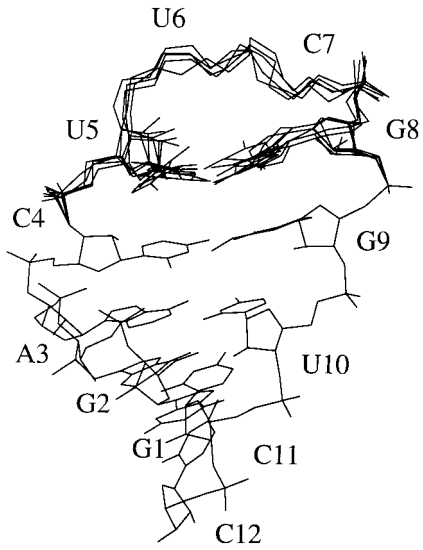


**Figure 3.** A drawing of the RNA tetraloop system in which five LES copies were used for the entire UUCG loop. All five copies are attached to the (single) stem. For clarity, hydrogen atoms, water, counterions, and U6 and C7 riboses and bases are not shown.

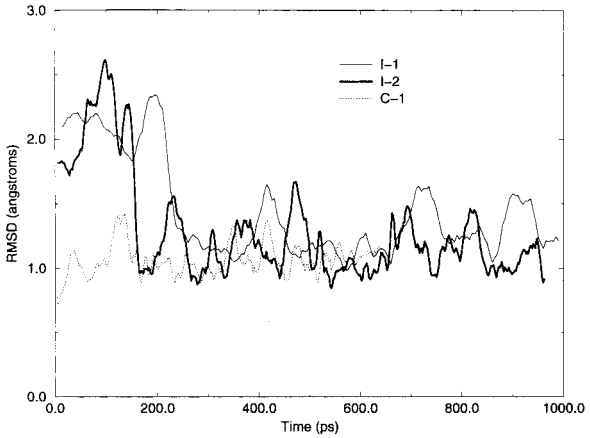


**Figure 4.** The loop RMSD (compared to the correct structure) as a function of time. Five LES copies were employed for the entire UUCG loop. Simulations I-1 and I-2 (solid lines) were initiated from the incorrect structure and at about 200 ps convert to a structure similar to that found when starting from the correct structure (simulation C-1, dashed line).

the same molecule (such as those with covalent bonds) do interact between image molecules. The requirement of calculating this correction and the larger number of particles in the system result in calculations that are somewhat more computationally intensive than the non-LES systems. The largest effect that was observed in the different applications of LES described below was simulations that were approximately 15% slower than the non-LES systems. However, the increase in efficiency of the LES simulations is much greater than the additional expense, as we describe in the Results section.

All of the simulations presented here used the same simulation parameters (other than LES) that were employed for the previous non-LES simulations for this system, including 11 neutralizing $Na^+$ counterions, ~2300 TIP3P water molecules and a standard all atom force field with fully charged RNA.[22] The time step was 2 fs, and SHAKE[23] was applied to all bonds involving hydrogen. Simulations were carried out in the NPT ensemble at a temperature of 300 K and pressure of 1 atm. The cutoff on Lennard-Jones interactions for PME simulations and all nonbonded interactions for non-PME simulations was 9 Å. The neighbor pairlist was residue-based and updated every 10 steps. For PME, a charge grid spacing of ~1 Å with cubic B-spline interpolation was used, and the direct sum tolerance was set to $10^{-5}$. Center of mass velocity was removed each 20 ps. All LES copies of individual atoms were initially assigned identical coordinates, moving apart after velocity reassignment at the start of the simulation. Simulations were carried out on a variety of parallel computers using different numbers of SGI R10000 CPUs. PME simulations required approximately 70% more computational effort than simulations using the nonbonded cutoff.

## Results

We first describe the various simulations that were carried out. These include changes in both the number of LES copies used and the regions that were chosen for enhanced sampling. We discuss the general behavior of these simulations, compare the results, and then provide a more detailed analysis of selected

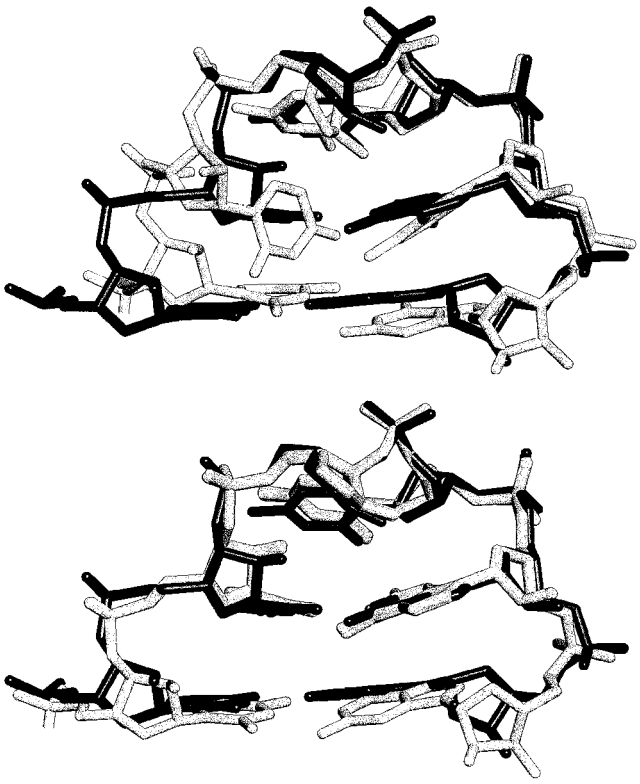

**Figure 5.** Comparison of the loop regions in the correct NMR structure (black) and the average LES MD structure for simulations using five LES copies of the UUCG loop region. Only the loop and the first base pair of the stem are shown (residues 4−9). The upper figure shows the initial (incorrect) structure. In addition to the base pair hydrogen bond differences shown in Figure 1, there is severe buckling of the U5:G8 base pair as well as other significant differences in backbone conformation on the 5′ end of the strand. The lower figure shows the same comparison after 500 ps of LES MD. Nearly all of the differences have been corrected.

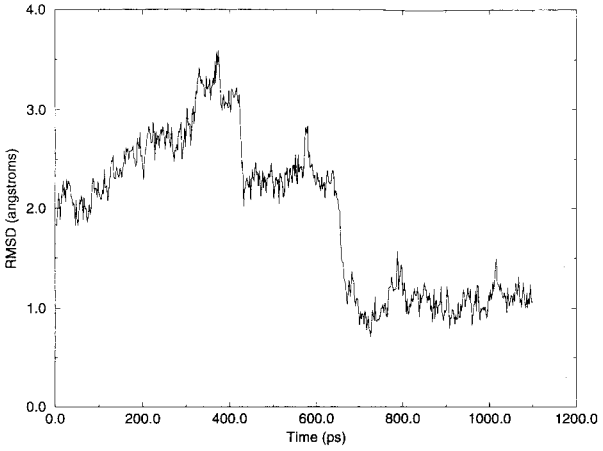


**Figure 6.** The loop RMSD (compared to the correct structure) as a function of time. Two LES copies were employed for the entire UUCG loop. This simulation also converts to the correct conformation, but on a slower time scale than when five LES copies were used.

simulations. The results of all simulations with RMSD values and the time of the conformational change (if appropriate) are summarized in Table 1. [All RMSD calculations include non-hydrogen atoms in the UUCG tetraloop (residues 5−8) except the base atoms of U6, which does not form specific contacts and shows higher mobility. This is also the variable position in the UNCG family of RNA tetraloops. For LES systems, the

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(23) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

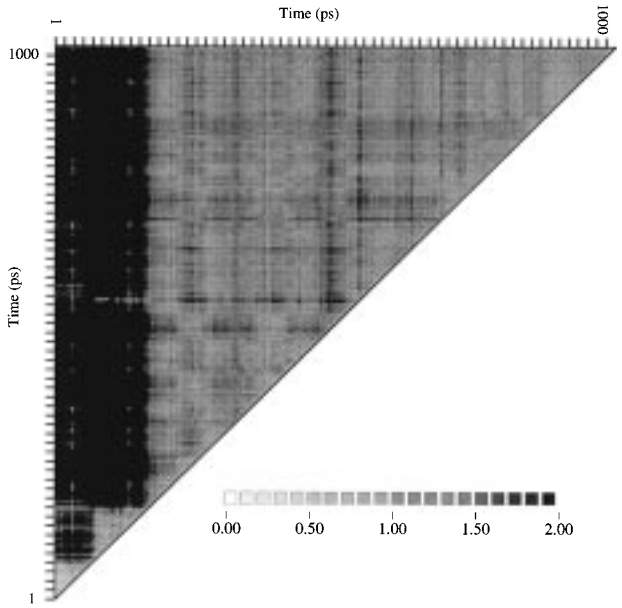


**Figure 7.** A 2-D RMSD plot, in which structures at different points during the simulation are compared to each other. The data are from simulation I-2, described in Figure 4. Values farther from the diagonal represent comparisons between structures farther apart in time, and light regions correspond to families of similar structures. In this simulation, three families are visited: the initial incorrect structure, an intermediate conformation (Figure 8), and the final (correct) conformation which persists for the last ~800 ps.

RMSD value is a comparison to a structure where all LES copies are in the correct conformation.]

It is not obvious how to decide which portions of a complex system require sampling enhancement. For example, because the chemical modification which succeeded in producing the conformational change involved converting the loop sugars from ribose to deoxyribose, and we wanted to make a minimal perturbation of the molecule, our initial application of LES involved replacing each of these 2′ hydroxyl hydrogen atoms with 5 LES copies. Previous studies found that five LES copies is a reasonable choice,[10] but we also explore the sensitivity of the LES/PME results to this number. The final structure after 2 ns of single copy MD on conformation I (which failed to convert to the correct structure)[19] was used as the initial structure, with all LES copies of atoms having identical initial coordinates. Simulations were then carried out for this system using either a cutoff on nonbonded interactions or the PME treatment for long-range electrostatic interactions.

Previous simulations of this RNA tetraloop using a nonbonded cutoff did not result in stable trajectories. Only when long-range electrostatics were included through PME were stable simulations obtained.[5] Since LES should improve conformational sampling, we do not expect LES with a cutoff to provide reasonable trajectories. To confirm that LES simulations also required PME, however, we carried out the appropriate control simulation for the LES system with a nonbonded cutoff. As shown in Figure 2, the RNA was unstable under these simulation conditions and within 100 ps underwent rapid conformational changes which involved loss of all base pairs. We therefore conclude that PME is also critical with LES and employed it for all remaining simulations.

During the LES simulation with PME, the copies of the 2′ hydroxyl hydrogens moved apart and did undergo more frequent transitions between rotamers as compared to the non-LES simulations. However, the energetic barriers between conformers I and C were still too high as there was no change in the loop conformation during 2 ns of MD, as shown in Figure 2.

Next, we made the most general choice and used five LES copies of the *entire* UUCG loop (Figure 3). Each of these copies was attached to the stem, and the stem interacted with these copies in an average way. We again initiated the simulation beginning from the incorrect structure, with all LES copies having identical initial coordinates. In Figure 4 we show the RMSD of the loop atoms (compared to the correct structure) as a function of simulation time. A reduction in RMSD at 175 ps to ~0.7 Å demonstrates that the simulation converted from the incorrect conformation to that obtained when starting from the correct structure and was successful in achieving the transition from conformation I to C for all five copies (Figure 5). Further details of this transition and the final structure are discussed below.

To explore the robustness of the method, two additional simulations were carried out for the same LES system. In simulation 2, the same conversion of I to C was observed for all LES copies using slightly different initial conditions (Figure 4). In this case, a single transition was observed at ~200 ps, with a final RMSD value of 1.0 Å compared to the MD average correct NMR structure and 0.5 Å compared to the average structure (after transition) from simulation 1. The third LES simulation was initiated with all five copies in structure C after equilibration using a single copy.[19] In this case we observe no significant changes during 750 ps of MD (Figure 4). These results demonstrate that the LES simulations remain in the correct conformation when placed there and are able to reproducibly locate this conformation from the incorrect NMR conformation.

We next tested the sensitivity of the results to the exact number of LES copies used. We created a new system which employed two, rather than five, LES copies for the UUCG loop. We again initiated the simulation in the incorrect conformation. Once again, we observed the transition from incorrect to correct conformation (Figure 6). Since the barriers to conformational transition with two copies are not reduced to the same extent as when five copies are employed, we expect and observe that this transition takes place on a longer time scale (~600 ps) than either of the simulations with five copies. However, two copies are still sufficient to improve the sampling and result in the I → C transition that was not observed during single copy simulations.

We further examined the sensitivity of the results to the size of the region chosen for application of LES by copying a larger portion of the RNA sequence. In this case, we used five copies of the central six nucleotides (CUUCGG), including the CG base pair at the top of the stem (Figure 1) in addition to the loop residues that were copied previously. Once again, each of these copies was attached to the neighboring non-LES stem regions, which interacted with these copies in an average way. The initial coordinates were the incorrect structure, and all LES copies had identical coordinates. In this simulation we again observe the I → C transition, in this case within 100 ps. The final structure of the loop differed from that obtained using the smaller LES region by only 0.7 Å.

As described above, changing the loop riboses to deoxyribose did result in observation of the I → C transition during MD at 300 K. However, several torsion values in the G8pG9 step did not convert to values that are observed in the structures obtained from refinement of the NMR data. These include the G9 $\alpha$, $\beta$, and $\gamma$, each of which has two different values represented in the family of NMR structures. These torsion differences are

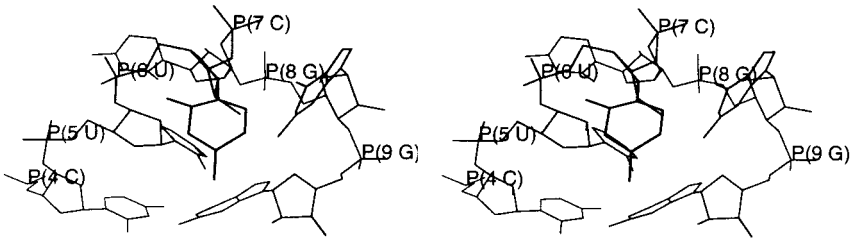


**Figure 8.** A stereo diagram of the intermediate conformation from the simulation described in Figure 7. The U5:G8 base pair hydrogen bonds have been lost and the G8 base is exposed to solvent. For clarity, four of the five LES copies, the first three base pairs of the stem, water, and counterions are not shown.

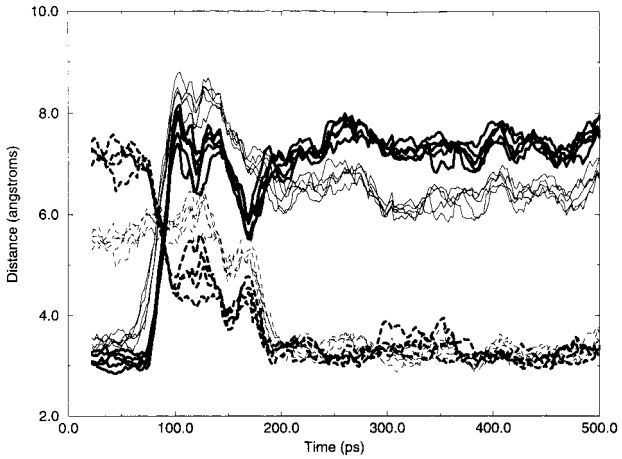


**Figure 9.** Distances between heavy atoms corresponding to the hydrogen bonds in the U5:G8 base pair shown in Figure 1. The simulation is initiated with the incorrect pattern and converts to the correct one. Line styles are the same as shown in Figure 1, and the values are running averages over 20 ps. Five LES copies were employed, and the hydrogen bond distances are shown for each copy.

**Table 2.** Torsion Values for the Three Dihedrals in G9 That Did Not Convert to the Correct Values in the Simulations That Used DNA in the Loop Region[a]

| | NMR[18] | DNA loop[19] | LES UUCG | LES CUUCGG |
|---|---|---|---|---|
| G9 $\alpha$ | −118 | −65 | −71 | −116 |
| | 112 | | | 127[b] |
| G9 $\beta$ | 92 | 158 | −162 | 86 |
| | −124 | | | −155[b] |
| G9 $\gamma$ | −168 | 70 | 69 | 81 |
| | 99 | | | −170[b] |

[a] When five LES copies were used for the same region, the same incorrect values were retained. However, when the LES region was extended to include G9, two families are structures are populated during MD, with values similar to those observed in the two families of refined NMR structures. [b] Largest population.

compensatory and do not significantly change the structure of the loop or stem regions. The values of these torsions in the NMR and simulated structures are presented in Table 2. We examined the values from the LES simulations where additional copies of only residues 5−8 (UUCG) were employed. In these simulations we observe that the initial incorrect torsion values are maintained even after the loop converted to the correct conformation. This behavior is similar to that observed in the DNA simulations and is likely due to the lack of LES copies in this region.

We next examined the values for these torsions obtained from the simulation where the LES region was extended one base pair into the stem on each side (CUUCGG). The LES copies now improve sampling for the G9 backbone, allowing it to more easily overcome barriers to conformational transition. Since LES does not necessarily provide accurate Boltzmann-weighted

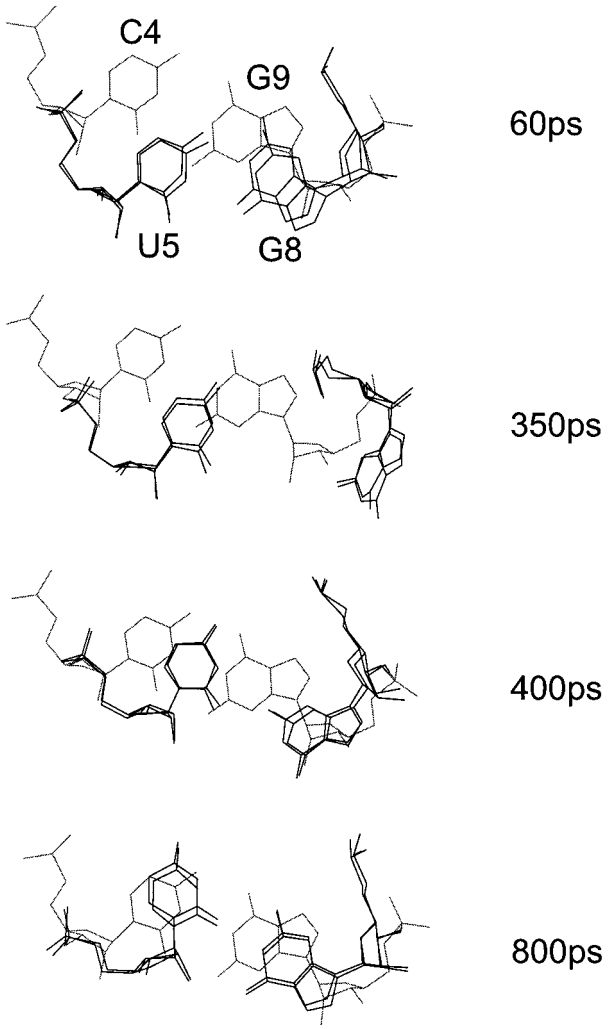


**Figure 10.** Snapshots of the U5:G8 base pair during a simulation that employed two LES copies of the UUCG loop. Also shown is the C4:G9 base pair in the stem which has a stacking interaction with the U5:G8 pair. For clarity, the rest of the system is not drawn. At 350 ps, the bases separate, and re-form at 400 ps with partially correct hydrogen bonds but with stacking against the stem that is similar to the incorrect conformation. At 800 ps the correct hydrogen bond pattern and stacking have been achieved. This conformation is retained throughout the remainder of the simulation.

populations, and the NMR data in this region is only semi-quantitative, we only compare the general features of the values obtained. For each of the three torsions, two different ranges of values are sampled during the simulation, in accord with the NMR data (Table 2). These data show that the LES simulation not only converts from initial incorrect torsion values to those observed experimentally but also demonstrates that more than one conformation may be accessible.

A major difference between the two experimental structures

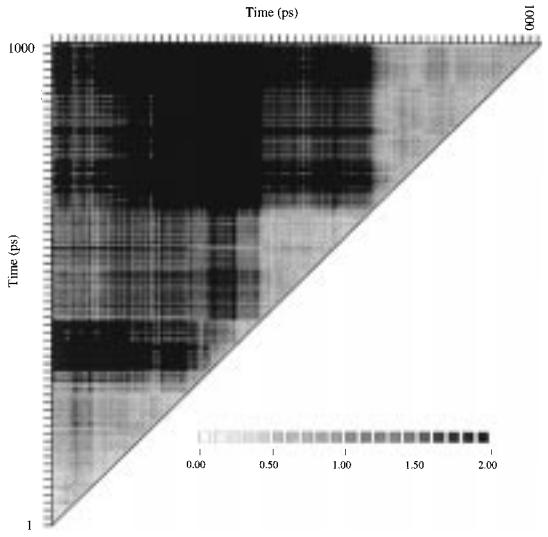


**Figure 11.** A 2D-RMSD plot for the simulation in which two LES copies were employed for the UUCG loop region. In this case four different families are visited (Figure 10). The light areas off-diagonal correspond to greater similarity between families of structures that are separated in time For example, the first and third families, which have similar stacking patterns for the U5:G8 and C4:G9 base pairs (Figure 10), are most similar.

is the hydrogen bond pattern between the bases in residues U5 and G8 (Figure 1). In Figure 7 we show the 2-D root-mean-square deviation (RMSD) of the loop atoms during a LES simulation using five copies of the UUCG loop region. Three families of structures (i.e., substates, indicated by light colored areas of the plot) are observed: at approximately 75 ps there is a conversion to a new structure (that differs by ∼2 Å). In this conformation (shown in Figure 8), the U5-G8 base pair has been separated, and the G8 base is flipped out, exposing it to the solvent. At ∼175 ps another transition occurs to the third conformer, in which the U5-G8 base pair re-forms and persists throughout the remainder of the 1 ns simulation. In Figure 9 we show the heavy atom distances corresponding to these hydrogen bonds as a function of time during the same simulation The breaking of the N3−O6 and O4−N1 reverse-wobble hydrogen bonds and formation of the bifurcated pattern involving O2−N1 and O2−N2 again demonstrate the conversion of all five copies from I to C, the correct structure. As shown in the plot, this switch represents significant changes which involve motions of up to 5 Å.

Observation of a single transition pathway in a simulation does not mean that this is the only pathway that exists between the two conformations. We therefore carried out a similar analysis for the alternate simulation that employed five copies of the UUCG loop and found that the transition in this simulation followed a different pathway. In this case only two conformations are visited, I and C, with no observation of an intermediate conformation involving significant solvent exposure of the G8 base. This result further demonstrates the importance of affordable simulations of conformational transitions: most current simulation methodologies do not allow for even a single observation of these events and therefore cannot provide insight concerning the existence and nature of multiple transition pathways.

Since these transition events occurred on a relatively fast time scale with five LES copies, we also carried out a detailed analysis of the simulation with two LES copies of the same region. In the case, the transition is less rapid, and there are two distinct intermediate conformations (Figure 10). In the first, the G8 base is exposed to solvent in a structure similar to that observed with five copies. At about 400 ps, the UG base pair re-forms with the correct hydrogen bond pattern. In this case, however, the relative orientation of these bases and the stacking of the UG pair against the stem CG pair differs from that found in the correct conformation (Figure 11). This structure persists for ∼200 ps, when the UG pair shifts relative to the CG pair and attains the correct stacking pattern and relative orientation, which is retained throughout the remaining 450 ps. These two changes occur as a single event when five copies are employed, most likely due to the reduction of the energy barrier between the two different stacking patterns.

## Conclusions

These simulations demonstrate that (1) the LES system remains in the correct location in phase space (when placed there), (2) the LES system can reproducibly find this location from an alternate conformation in ∼200 ps, and (3) while the amount of time required for the transition varies somewhat, the overall results are not very sensitive to either the number of copies employed or the size of the LES region, as long as the UUCG loop was copied. Several independent single copy MD simulations (totaling 5.5 ns) starting from I showed *no conversion* to the C conformation. The reduction in barrier heights provided by LES therefore results in simulations that are more than an order of magnitude more efficient than single copy methods. However, these results are only one example, and we plan to test the general applicability of the method using other protein and nucleic acid systems. We believe that this approach is also likely to be an important component in theoretical predictions of protein structure from amino acid sequence, where both extensive conformational sampling and the accuracy of the energy function are of critical importance.

**Acknowledgment.** P.A.K. acknowledges research support from NIH-CA-25644. C.S. is grateful to Tom Darden for useful discussions.

JA9727023