Improved Conformational Sampling through an Efficient Combination of Mean-Field Simulation Approaches

Xiaolin Cheng,† Viktor Hornak,‡ and Carlos Simmerling*,†,‡

Department of Chemistry and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794-3400

Received: February 27, 2003; In Final Form: July 1, 2003

In this article, we show for the first time the combination of locally enhanced sampling (LES) and the generalized Born continuum solvation model. It has been shown that the reduction in barrier heights provided by LES can result in solvated simulations that are more than an order of magnitude more efficient than single-copy methods, but we note that explicit solvent introduces a strong correlation between copies, reducing the benefits of the technique. The use of continuum solvation can also accelerate conformational transitions because of reduction in friction. We describe a direct combination of the two algorithms that maintains the equivalence of the LES and non-LES global energy minima and eliminates the solvent-induced coupling of the copies. Since this approach results in a significant increase in computational requirements, we also present reasonable approximations that greatly increase efficiency. We have applied the resulting combination to simulation of conformational change in an RNA UUCG tetraloop and have shown that the combined GB + LES approach is more efficient than use of either GB or LES alone. We carried out a large number of these simulations and show in a converged manner that the rate constant for the conformational transition is significantly increased with GB + LES as compared to GB alone. In addition, we demonstrate that the combined method significantly improves the ability of LES copies to explore independent transition paths as compared to LES simulations with explicit solvation. We thus believe that this GB + LES technique may be a useful component of structure refinement and prediction studies.

Introduction

The quality of results obtained from a molecular simulation depends critically on two factors: the energy function must provide an accurate model of the underlying physics of the system and the simulation should adequately sample the important regions of the resulting energy landscape. These issues are coupled; a more accurate energy function may be more difficult to adequately sample. A compromise between accuracy and sampling must therefore be reached, and approximations to the true energy function may be unavoidable in order to improve sampling. For biomolecular systems, classical all-atom models with empirical parameters are typically used to obtain more complete sampling than could be obtained with a more accurate (e.g., fully quantum mechanical) description.[1−4] Even when classical force fields are employed, obtaining adequate sampling during simulations of large, flexible systems on a computationally affordable time scale is a major obstacle to more extensive application of these tools.

The choice of solvation models also influences both accuracy and sampling. An accurate description of solvation is critical to the success of modeling biological systems.[5] Explicit inclusion of solvent molecules[6,7] has proven very successful for biomolecular simulations, particularly[8] when combined with efficient approaches to treat long-range electrostatics, such as particle mesh Ewald (PME).[9,10]

While explicit solvation may provide an atomic-detail model of solvation, the cost associated with computing forces and integrating equations of motion for the large number of explicit solvent atoms reduces the number of solute conformations that can be evaluated. In addition, frictional and packing effects from solvent may result in a slower time scale for the process of interest, thus requiring calculation of an increased number of time steps to model events of interest using molecular dynamics simulations. Thus, explicit solvation can significantly limit conformational sampling, and obtaining well-converged sampling in explicit solvent is still far from trivial.[11−13]

An alternative approach to modeling the electrostatic effects of solvation is through a continuum description, such as the generalized Born (GB) model.[14−16] One of the key advantages of GB over an explicit solvent model is that it is much more computationally efficient. Only the solute degrees of freedom are considered explicitly, and solvent is approximated as a dielectric medium that influences the behavior of the solute atoms. Furthermore, it has been shown that convergence of biomolecular simulations is accelerated with frictionless implicit solvent models, so that in many cases fewer simulation steps are needed to model a particular transition (as compared to an explicit model).

In simulations reported by Tsui and Case,[17] duplex A-form DNA (d(CCAACGTTGG)$_2$) converged to B-form more than 20 times faster in GB than in explicit solvent. Williams and Hall studied[18] the applicability of the GB model to an RNA tetraloop system, for which an important structural transition did not occur in standard MD simulations in explicit solvent.[19] With GB solvation, the structural transition became accessible on the ns time scale in otherwise standard MD. These studies strongly imply that molecular dynamics simulation with the GB model can explore phase space much more efficiently than MD

* To whom correspondence should be addressed. E-mail: carlos.simmerling@stonybrook.edu.
† Department of Chemistry.
‡ Center for Structural Biology.

10.1021/jp034505y CCC: $27.50
Published on Web 12/05/2003

with explicit solvent. An additional "bonus" is that the cost per unit simulation time is often reduced because of the smaller system size. Thus, the simulations using GB may extend the effective time scale of processes that we are able to model and permit observation of events that are inaccessible or unaffordable in simulations with explicit solvation, such as the folding of small proteins.[20]

While a continuum solvent model may improve sampling in some cases, many barriers to conformational transitions do not arise from the solvent. Locally enhanced sampling[21−23] (LES) is a mean-field approach that has proven useful in improving sampling through a reduction in internal barrier heights. LES is effective even when an explicit solvent model is employed.[24,25] The details of the LES approach have been described in detail in the past and further detail is given below. In brief, the LES method provides the opportunity to focus computational resources on the portion of the system of interest by replacing it with multiple copies. The mean-field effect obtained from averaging the interactions among LES copies provides a smoothing effect of the energy landscape,[23,26] improving sampling efficiency through reduction of barrier heights. We recently showed that LES could be efficiently used to optimize conformations of proteins loops,[27] although in that case we employed a distance-dependent dielectric solvation model since we had not completed the GB + LES method described here.

Since a major benefit of LES is the ability to simultaneously obtain multiple trajectories for the copied portion of the system, it is desirable to maximize the independence of the replicas during the simulation to increase both the amount of phase space that is sampled and the magnitude of the mean-field smoothing effect. While a major advantage of LES is that it can be successfully employed with an explicit solvent model,[24,25] we have observed that the positional variance of solvent-exposed copies is not nearly as large as that obtained during similar simulations in the gas phase. This is likely due to the simultaneous interaction of solvent molecules with all of the copies; copy divergence therefore requires the creation of a larger solvent cavity. This results in a free-energy penalty analogous to the hydrophobic effect and tends to reduce the independence of the copies through an indirect coupling. In addition, the solvent molecules surrounding the group of copies may not be able to simultaneously provide ideal solvation for each of the copies. This issue is discussed in greater detail below.

Since LES and continuum solvation each increase effective transition rates (but through different approaches), we expect that overall sampling with the combined method should exceed that obtained when LES or continuum solvent are employed alone. This combination also permits an approach to solvation that provides greater independence and improved solvation for solvent-exposed copies. We therefore developed a combined GB + LES approach and implemented it in the AMBER suite of programs.[28]

We previously reported[29] that non-LES simulations for a pyrene-substituted DNA system using GB converged to either of two low-energy conformations more rapidly than we observed with explicit solvent. However, transitions between these structures (of similar energy) were not observed even with the continuum solvent. When we employed LES in explicit solvent, interconversions were seen but the simulations were much more computationally demanding because of the explicit solvent. These observations provided additional incentive for the development of the combined GB + LES approach presented here.

As a model system to test this technique, we chose the RNA UUCG tetraloop ($G_1G_2A_3C_4[U_5U_6C_7G_8]G_9U_{10}C_{11}C_{12}$), for which structures have been determined by NMR.[30,31] This makes an excellent model because of its small size and since several previously reported theoretical studies explored the conversion of an incorrect conformation (I) for the loop region into the correct one (C). Standard MD simulation in explicit solvent resulted in *no* conversion of I to C in several ns of MD.[19] The use of LES to make multiple copies of the loop region in explicit solvent resulted in reproducible, spontaneous conversion of I to C in about 200 ps.[32] Single-copy GB simulations were also successful in the I→C conversion within about 1200 ps.[18] Thus, each of these approaches to enhanced sampling was successful, and we can compare these results to those from our combined GB + LES algorithm.

The GB + LES simulations that we present converge more rapidly than the single-copy GB or explicitly solvated LES simulations, suggesting that the sampling enhancements provided by these two approaches are complementary. Perhaps more important, however, is the observation that the copies in combined GB + LES are able to sample alternate transition pathways in a single simulation, which was never observed with LES in explicit solvent and is (of course) not possible with standard MD simulations.

Methods

Locally Enhanced Sampling. In our current implementation of LES in AMBER (version 7), partial charges, Lennard-Jones parameters, bond and angle force constants, and dihedral barriers are all scaled such that the total system energy is equivalent to the average energy of the multiple non-LES "reference" systems. We define a reference system as the single-copy system obtained by combining all of the atoms belonging to one copy with the noncopied atoms. Thus, there are N reference systems for an N-copy LES system. We will refer to these hypothetical reference systems at several points in this document, and a specific example for a simple model system is described below. Different copies do not interact with each other during the simulation and interact with the noncopied atoms in an average way. Calculation of energies and forces for this system can be more efficient than the corresponding calculations for all of the separated reference conformations, since interactions involving the noncopied region are only calculated once for all reference conformations when LES is used.

It can be shown[33] that the global energy minimum of the LES system occurs when all copies occupy the position of the global minimum of the non-LES system. This is an extremely valuable property of LES, particularly for structure optimization, and our goal is to maintain this correspondence in our combined GB + LES approach. As described above, the energy function for the LES system is constructed such that the potential energy is the arithmetic average of the energies of the reference systems. As a result, the energies of each copy do not depend on the coordinates of the other copies. Since the copy energies are independent, the global energy minimum of the LES system must occur when each copy is in its own global minimum. In other words, if moving one copy to an alternate location results in lower system energy, then moving the others to the same location would also reduce the energy. The global minimum is therefore a configuration with all copies in identical positions. Since the LES energy is an average of the corresponding single-copy energies, any LES configuration in which the copies have the same coordinates must have the same energy as the corresponding non-LES system. As described above, this is the criterion for constructing the LES energy function. Therefore, the LES global energy minimum is the non-LES configuration

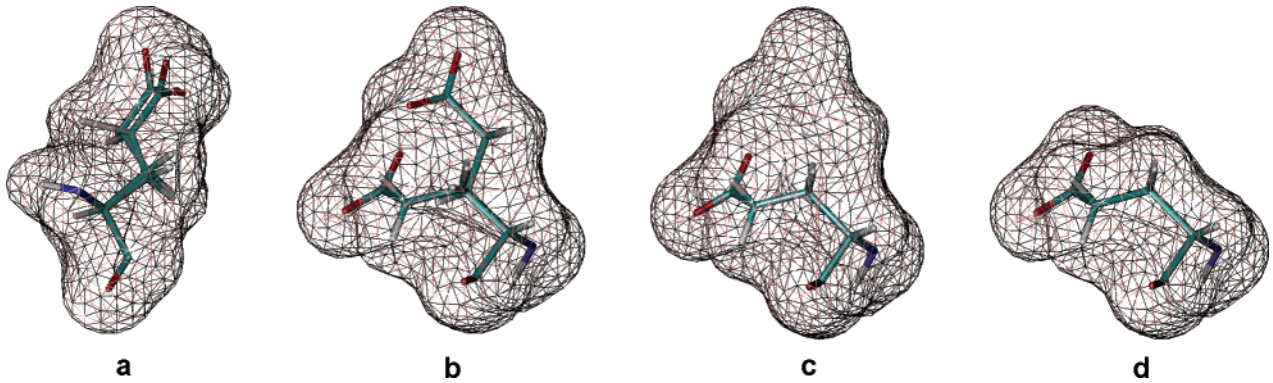

Figure 1. A Glu amino acid, shown with solvent-accessible surface area (SASA). (A) Two LES copies of the side chain *in explicit solvent* (solvent not shown), after equilibration with MD. The penalty associated with creating a larger solvent cavity typically results in only small variance in copy positions. (B**)** A snapshot of the same system in which the copies are sampling different conformations. The SASA represents the actual solvent cavity when the solvent interacts with both copies. (C**)** The non-LES reference system corresponding to one of the copies in Figure 1B. The SASA is that seen by the copy during LES simulation in explicit solvent, but the larger cavity represents a low-probability solvent configuration at normal T and P. (D). The desired reference system from Figure 1C, with correct SASA. The solvent configuration clearly differs from that sampled by the LES system.

with the lowest energy, that is, the non-LES global energy minimum. Other local minima on the original (non-LES) energy landscape have corresponding minima for LES (with all of the copies in the position of the original minimum), but the LES landscape also introduces many new local minima in which copies simultaneously populate different local minima from the non-LES landscape.

An advantage to LES compared to many approaches to improved sampling is that it can be employed with explicit inclusion of solvent molecules.[32] However, when solvent-exposed copies occupy similar but nonidentical positions, explicit solvent is excluded from a volume corresponding to the region occupied by any of the copies. As a result, none of the copies samples a truly solvent-exposed state (Figure 1). In this case, LES may give efficient sampling of states of the reference system that are possible, but of lower probability than those with the copies in direct contact with solvent. This effect is inversely correlated to that described above; as the copies become more independent, the instantaneous solvation becomes less representative of what would be expected for a non-LES system exploring the same conformations. In the limit of 0 K, the problem may disappear if the copies converge to identical positions, but at finite temperatures the thermal fluctuations typically result in varying degrees of this undesirable behavior.

Thus, the LES system with explicit solvent corresponds to a set of reference systems that are not ideally solvated. If LES is combined with a continuum solvent model, it is possible to independently solvate each of the reference systems. Our strategy in deriving this combination is to maintain the correspondence of the LES system energy and the average energy of the reference systems, while providing a more realistic representation of solvation for the individual copies.

Generalized Born Solvation. In the present study, we will combine the LES approach with the GB method for calculation of the electrostatic component of solvation free energy. Under the GB approximation,[14] this is given by eq 1

$$\Delta G_{\mathrm{pol}}{}^{\mathrm{GB}} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{ij}\frac{q_i q_j}{f_{\mathrm{GB}}} \qquad (1)$$

A Debye−Huckel term is incorporated[34] into AMBER to account for salt effects at low salt concentration; eq 1 thus becomes

$$\Delta G_{\mathrm{pol}}{}^{\mathrm{GB}} = -\frac{1}{2}\left(1 - \frac{e^{-\kappa f_{\mathrm{GB}}}}{\epsilon}\right)\sum_{ij}\frac{q_i q_j}{f_{\mathrm{GB}}} \qquad (2)$$

where q_i and q_j are atomic partial charges, ϵ is the solvent dielectric constant, κ is the Debye−Huckel screening parameter, and the double sum runs over all pairs of atoms. f_{GB} depends on the effective Born radii α_i, α_j, and the distance r_{ij} between atoms:

$$f_{\mathrm{GB}} = \left[r_{ij}{}^2 + \alpha_i\alpha_j \exp\left(-\frac{r_{ij}{}^2}{4\alpha_i\alpha_j}\right)\right]^{1/2} \qquad (3)$$

As $r_{ij} \rightarrow 0$, $f_{GB} \rightarrow \alpha_i$, the effective radius that establishes the self-energy of charges that arises from polarization of the surrounding dielectric medium.

GB + LES: Difficulties with the Effective Born Radii. The main challenge in combining GB and LES arises when calculating the effective Born radii for each atom. In AMBER (without LES), the effective Born radii α_i are calculated via the pairwise descreening approximation,[35]

$$\alpha_i{}^{-1} = \rho_i{}^{-1} - \frac{1}{2}\sum_j \int_{L_{ij}}^{U_{ij}} dr\left(\frac{1}{r^2} - \frac{r_{ij}}{2r^3} - \frac{1}{2r_{ij}r} + \frac{S_{ij}{}^2\rho_j{}^2}{2r_{ij}r^3}\right) =$$

$$\rho_i{}^{-1} - \frac{1}{2}\sum_j \left[\frac{1}{L_{ij}} - \frac{1}{U_{ij}} + \frac{r_{ij}}{4}\left(\frac{1}{U_{ij}{}^2} - \frac{1}{L_{ij}{}^2}\right) + \frac{1}{2r_{ij}}\ln\frac{L_{ij}}{U_{ij}} + \frac{S_{ij}{}^2\rho_j{}^2}{4r_{ij}}\left(\frac{1}{L_{ij}{}^2} - \frac{1}{U_{ij}{}^2}\right)\right] \qquad (4)$$

where

$$L_{ij} = \begin{cases} 1 & \text{if } r_{ij} + s_{ij}\rho_j \le \rho_i \\ \rho_i & \text{if } r_{ij} - s_{ij}\rho_j \le \rho_i < r_{ij} + s_{ij}\rho_j \\ r_{ij} - s_{ij}\rho_j & \text{if } \rho_i < r_{ij} - s_{ij}\rho_j \end{cases}$$

and

$$U_{ij} = \begin{cases} 1 & \text{if } r_{ij} + s_{ij}\rho_j \le \rho_i \\ r_{ij} + s_{ij}\rho_{ij} & \text{if } \rho_i < r_{ij} + s_{ij}\rho_j \end{cases}$$

In GB, the effective radius of an atom, and therefore the interaction between any pair of atoms, is no longer independent of the coordinates of the rest of the system because of the descreening effects of the other atoms. This is the cause of problems employing the GB model with frozen atoms.[36] If we take the approach that each noncopied atom should be simultaneously descreened by *all* of the LES copies, the corresponding reference systems would have the individual copies occupying a solvation shell corresponding to the space occupied by all copies—directly analogous to the situation with LES in explicit solvent. Each reference solute would not be fully solvated (Figure 1) and would therefore represent less probable, though possible, configurations of the reference systems. This is not the ideal average as this partial desolvation of the copies is one of the problems encountered with LES and explicit solvent that we wished to avoid. Since different LES copies can also occupy the same space, the pairwise descreening approach[35] that is used in AMBER to calculate effective Born radii would not be reasonable since the descreening effects due to the multiple copies would not necessarily be additive because of allowed overlap between different copies.

An additional problem with this approach is that moving any copy would potentially change the effective radii of all atoms and therefore directly affect the energetics of interactions in other copies. Thus, our proof of equivalence of global minima presented above would no longer remain valid. This problem does not arise with explicit solvation, since the nonbonded energy can still be separated into a sum of pairwise terms and thus copy independence is maintained. Even if an alternate formalism were employed in which this coupling was not present, the effective solvation of the system would incorporate the same inaccuracies as encountered with explicit solvent, that is, the effective solvation cavity would surround the set of copies rather than represent individual solvation of each reference system.

We therefore take the approach of explicitly enforcing the correspondence with the average of the reference systems that the LES system represents, including correct reference solvation of each copy. We first imagine separating the copies, combining each with the noncopied region, and calculating effective radii in each resulting system. In this case, every atom in each reference system will require a unique effective radius as compared to the same atom in the other systems: the LES copies of an atom can occupy different positions in space and therefore need different radii, but the non-LES atoms also require multiple radii, one for each of the hypothetical systems, since the descreening of these atoms will differ among reference systems because of the changes in the positions of LES atoms. In the actual simulation, we do not explicitly separate these hypothetical systems, but the interactions involving any pair of non-LES atoms are no longer identical in the reference systems, and obtaining the correct average over the reference systems requires explicit calculation of these interactions multiple times using each of the sets of effective radii.

To clarify this issue, we describe a simple model system with four atoms, A, B, C, and D (Figure 2). We replace atoms C and D with two copies, C^1/C^2 and D^1/D^2 where the superscript denotes the copy number. Since different copies of a region do not interact, C^1 does not interact with C^2 or D^2, and so on. The energy of this system is an average of the energies of the two reference conformations represented by LES: ABC^1D^1 and ABC^2D^2. In standard LES, the interaction between A and B does not depend on the coordinates of C and D, and this interaction is factored out of the average and calculated only once. Thus, the form of the LES averaging in this case is given by the following equations:

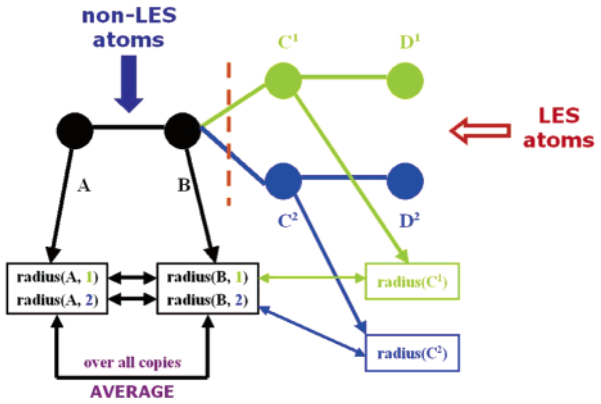

Figure 2. A simple model system to clarify the calculation and use of effective radii. In a system with four atoms A, B, C, and D, atoms C and D are replaced with two LES copies. The radii are described in the text.

$$V_{\text{real}} = V(\text{A,B,C,D}) = V_{\text{non-LES}}(\text{A,B}) + V_{\text{non-LES,LES}}(\text{A,B,C,D}) + V_{\text{LES}}(\text{C,D})$$

$$V_{\text{LES}} = V_{\text{non-LES}}(\text{A,B}) + \frac{1}{2}[V_{\text{non-LES,LES}}(\text{A,B,C}^1,\text{D}^1) + V_{\text{LES}}(\text{C}^1,\text{D}^1) + V_{\text{non-LES,LES}}(\text{A,B,C}^2,\text{D}^2) + V_{\text{LES}}(\text{C}^2,\text{D}^2)]$$

where

$$V = \Sigma\,(E_{\text{bond}} + E_{\text{angle}} + E_{\text{tor}} + E_{\text{ele}} + E_{\text{vdw}} + E_{\text{GB}})$$

The first step in a GB calculation is to determine effective Born radii for all atoms. As stated above, both LES and non-LES atoms require a separate effective radius for each of the reference systems. For example, radii radius(A,1), radius(B,1), radius(C,1), and radius(D,1) are calculated within reference conformation 1 $\{ABC^1D^1\}$ and radius(A,2), radius(B,2), radius(C,2), and radius(D,2) are calculated within conformation 2 $\{ABC^2D^2\}$. For this calculation, the increase in effort required is smaller than the number of copies since some duplicate calculations can be avoided. For example, the descreening contribution of atom A because of B is the same in both systems because all of the systems have this same atom pair at the same separation distance. For small LES regions, this can result in large improvements to the efficiency of calculation of radii (analogous to factoring of duplicated nonbonded interactions in traditional LES). Not all descreening contributions are the same; the descreening of atom A because of C differs in the two systems since C has different coordinates.

Calculation of the atomic forces due to the GB solvation with LES proceeds in an analogous manner to non-LES GB. For interactions between an atom in LES copy *i* and a non-LES atom, the energy and forces are calculated using the single radius for the LES atom and radius *i* for the non-LES atom. Only one calculation is performed, and the difference due to LES is just the selection of the radius for the non-LES atom that was calculated using the descreening from the corresponding LES copy. For example, the interaction between A and C^1 is calculated using the first copy of the effective radius for atom A, since atom C^1 belongs to LES copy 1. For the interaction of A and B, we calculate the interaction twice: once with radii set 1 and again with radii set 2, each corresponding to alternate conformations of the LES region that affect the screening of

this interaction. The computational effort required for this part of the calculation (the pairwise electrostatics) increases by a factor of the number of LES copies employed. This is in contrast to the calculation of the multiple sets of effective radii, which may not require significantly more effort since most of the contributions to the descreening were shared among the sets.

In the limit of a very small LES region, the number of effective radii and nonbond interactions that need to be calculated will increase by a factor of the number of reference conformations represented by the LES copies. This differs dramatically from the non-GB LES approach we implemented in AMBER, in which interactions in the noncopied region were identical among all of the reference systems and were therefore factored out of the average and calculated only once. This factoring of duplicated terms (along with the mean-field effect) is the source of the computational efficiency of using LES as compared to multiple simulations of the entire system, especially when the LES region is a small fraction of the system.

This procedure provides an exact average of the energies and forces of the separated and properly solvated reference systems but is less computationally efficient. However, the mean-field effect of LES is still obtained, and dynamics on the smoother energy landscape may provide greater efficiency than multiple individual trajectories. In addition, we present reasonable approximations below that avoid the majority of this computational overhead.

Approximations. The forces and energies calculated using the approach described above represent an exact average of the reference systems, and no approximations have been introduced beyond those already inherent in LES and GB. In this case,the pairwise interactions using GB + LES with N copies take N times as long as a corresponding non-LES calculation. For each pairwise interaction in the original system, there exist N times as many interactions in the LES system: N interactions involving the N copies of each LES atom, and N components in the average for each fully non-LES pair. The calculation of effective Born radii can be simplified, since the fully non-LES pairwise descreening contributions are identical in all reference systems and can be shared when calculating the multiple radii of the non-LES atoms. Other drawbacks to the introduction of multiple effective radii include larger memory requirements and increased communication overhead in parallel implementations, especially important when using PC clusters with low network bandwidth. We therefore investigated approximations that would retain a high level of accuracy while reducing the number of effective radii that were required.

The source of the need for multiple radii for non-LES atoms is that the descreening effects arising from the LES region may differ for different copies, since the copy conformations may differ. When the atoms are far from the LES region, this effect is reduced, and the variation in effective Born radii due to the change in copy conformation is usually very small. This suggests that we can ignore the differential descreening effects of copied atoms at long distances, and the calculation of the radii would thus be faster.

This type of neglect of the effect of conformational changes for atoms at long distances is the essence of the reported[36] combination of GB with frozen atom approximation, in which atoms far from the moving region do not have their radii updated each step. In the present case, however, all of the atoms are moving and thus need to have their radii recalculated regardless of whether LES is used. We therefore do not use a distance cutoff for calculation of the effective radii. It is important to keep in mind where the extra work is needed: not in the calculation of the effective radii (since most terms involve fully non-LES pairs and therefore introduce no extra effort as compared to standard GB), but in *using* them. The key challenge is to reduce the number of non-LES atoms that need multiple radii to avoid the explicit enumeration of all elements in the average interaction with other non-LES atoms.

TABLE 1: Comparison of the Effects of Various RDT Values

RDT (Å)	N_{RDT}/N_{tot}[a]	ΔE[b] (kcal/mol)	RMSDf[c] (kcal/mol Å)	ΔF_{max}[d] (kcal/mol Å)	speedup[e]
0.1	0.99	0.70	0.0069	0.131	2.89
0.01	0.94	0.12	0.0013	0.022	2.46
0.001	0.78	0.03	0.0002	0.003	1.71

[a] N_{RDT}/N_{tot} indicates the fraction of non-LES atoms that can use a single effective radius instead of 5. [b] ΔE is the unsigned energy difference between calculations with/without RDT. [c] RMSDf is the root-mean-square deviation of forces between calculations with/without RDT; [d] ΔF_{max} is the maximum difference in atomic force components between calculations with/without RDT. [e] Speedup is the time required for evaluation of nonbonded energy and forces without RDT divided by the time with RDT.

We thus introduce a cutoff for the permissible deviation among the multiple radii of a given atom. We approximate that if the radii are similar, the average of the N interactions using the sets of radii can be approximated by one of the elements in the average. Therefore, in our implementation a threshold value was introduced to reduce the calculation time. When the differences among the N copies of effective radii of a non-LES atom are less than this threshold, it is reasonable to use a single effective radius. As a consequence, N pairs of calculations involving such atom pairs are reduced to one per pair, a factor of N speedup. Moreover, using one effective radius saves memory and reduces communication overhead since there are fewer effective radii to distribute among multiple computing nodes.

To test the feasibility of this approximation, we calculated the energy and force errors using various radii difference threshold (RDT) values. The test involves calculation of the atomic energies and forces for the LES system, and comparing these data to an exact value obtained from explicitly separating and averaging the N corresponding non-LES reference conformations. This test also indicates the relative computational efficiency compared to multiple non-LES simulations.

Since the non-LES region in our small RNA tetraloop model system is small, only a relatively small percentage of atoms are not copied and therefore most atoms are affected by differences in the conformations of the LES copies. The gain in efficiency due to the cutoff is therefore not as apparent as might be observed in a typical, larger system with a smaller LES fraction. We thus carried out the efficiency tests for triose phosphate isomerase (TIM), a protein composed of 247 residues. Five LES copies of loop 6 (the active site lid, residues 165−175) atoms were employed, and short GB + LES MD simulation was carried out to obtain a configuration with nonidentical coordinates for the LES atoms.

In Table 1 we provide the results of energy and force errors and relative performance using LES + GB, with various RDT values, for five copies of the active site loop in TIM. We observed that the energy and force errors are generally small enough for stable multi-ns MD simulation when the RDT is set to 0.01 Å (which means that the average effective radius for a non-LES atom is used if the difference among the multiple radii is less than 0.01 Å). In this case, ∼94% of the atoms use a single effective radius, and the calculation of nonbonded

interactions is nearly 3 times faster than without use of the RDT approximation (the speedup is not a factor of 5 since LES adds additional overhead besides that reduced by the RDT). Thus, it appears to be a reasonable approach, since LES is already an approximate method and the "exact" effective Born radii calculated through pairwise approximation are inherently imperfect and become the dominant source of error for solvation free-energy calculations.[37]

If one desires a more accurate trajectory, an RDT value of 0.001 Å results in an energy difference (compared to the exact calculation) of ~0.03 kcal/mol, and average atomic force deviations less than 0.0002 kcal/(mol Å). Even with this small RDT value, nearly 80% of non-LES atoms employ a single effective radius, and the calculation requires less than 60% of the time required without the RDT approximation.

Even with this approximation, the calculation using GB + LES is somewhat more computationally intensive and requires more memory than the non-LES calculations. However, we show below that the increase in efficiency of the LES simulations is much greater than this additional expense. The simulations not only converge more rapidly than corresponding non-LES simulations, but also show multiple transition pathways and time scales in single simulations, thus providing an improvement over non-LES approaches and LES simulations in explicit solvent.

Evaluating Convergence. The force fluctuation metric first introduced by Thirumalai[38−41] was used as a rigorous measure of the rate of sampling. Following their definition, the average force on the *i*th atom for the fluctuation metric is defined as

$$f_i^a(t) = \frac{1}{t}\int_0^t ds F_i^a(s)$$

where *a* indicates that the average is calculated over the *a*th trajectory. Given two independent trajectories, the time-dependent force metric can be defined as the difference between the averages calculated over the *N* frames of the pair of trajectories *a* and *b*:

$$d(t) = \frac{1}{N}\sum_{i=1}^{N} |f_i^a(t) - f_i^b(t)|^2 \qquad (5)$$

The AMBER program was modified to report these values during standard GB and GB + LES simulations.

Rates for the I→C transition were obtained by collecting first passage times for an ensemble of simulations initiated from the incorrect conformation. Random number seeds were varied among the simulations to provide different initial velocity distributions and divergent behavior. The time dependence of the fraction of the ensemble that remained in the incorrect conformation was fit to a single exponential, assuming first-order kinetic behavior. By collecting first passage times into the correct conformation, this procedure directly provides forward rate constants, rather than the sum of forward and reverse rate constants that are obtained from a traditional relaxation experiment. A loop region heavy atom RMSD value of 1.0 Å from the correct NMR structure was used as the threshold for determination of the transition event.

Simulation Details. The original incorrect and correct RNA tetraloop NMR models were used as starting structures.[30,31] The AMBER module ADDLES was used to construct the LES systems for simulation. All LES copies of individual atoms were initially assigned identical coordinates but unique velocities, and therefore diverged with propagation of time. The time step was 1 fs, and SHAKE was applied to all bonds involving hydrogen. No nonbonded cutoff was used. The AMBER ff94 force field[3] was used in all calculations with either the GB continuum model or a simple distance dependent (1/*r*) dielectric treatment, as noted when the simulation is described. The Born radii were adopted from Bondi with modification of hydrogen,[17] and the scaling factors for Born radii were taken from the tinker modeling package.[42] An offset of 0.09 Å was used for the radii.[14]

The crystal structure[43] of TIM (PDB code 1YPI) was used for single-point energy and force calculations. The same set of GB parameters was used for TIM as for the RNA tetraloop. To compare the effects of different RDT values on resulting forces and energies, we generated a TIM structure with five loop copies with loop region RMSD about 1.0 Å from each other. Of the 3778 atoms in the original TIM system, 150 loop atoms were replaced by five copies. The resulting LES system was composed of 3628 non-LES atoms and 750 LES atoms.

As a test of the program code, we investigated the accuracy of the average energies and forces calculated for the LES copies. First, we calculated energy and forces for a LES system in which the coordinates of each copy differed, and then divided the LES copies into the five reference (non-LES) systems, each using the same coordinates for the non-LES region. Energies and atomic forces were calculated for each system, and the averages of these values were compared to those obtained directly for the LES system. The values from each of these approaches to the averaging were identical, suggesting that the code was robust. This also confirms that the LES calculation is properly accounting for individual solvation of each copy, one of our main goals in the development of the algorithm.

Since the combination of GB and LES is nontrivial, we also tested the behavior of a distance-dependent dielectric treatment of solvation since this approach is both straightforward and efficient. However, simulations starting from incorrect and correct structures for the RNA tetraloop did not show behavior comparable to those observed in otherwise identical GB simulations. The correct structure was unstable and the incorrect structure did not convert to the correct structure. In both cases, RMSD values compared to the C conformation were ~3−4 Å, and the LES copies did not converge to a single conformation. These findings are consistent with similar instability for the tetraloop reported by Hall et al. for single-copy simulations with a distance-dependent dielectric.[18]

Results and Discussions

The major difference[30,31] between the two experimental structures of the RNA tetraloop is the hydrogen bond pattern between the bases in residues U5 and G8 (Figure 3). As described above, previous standard MD simulations of this RNA tetraloop in explicit solvent with PME did not result in interconversion between the two structures.[19] When LES was employed in explicit solvent, the correct conformation was stable, but the incorrect underwent a rapid transition to the correct form within 200 ps.[32] Transition from the incorrect to correct structure was also observed in about 1200 ps using the GB implicit solvent model without LES.[18] We therefore concluded that both GB model and LES sampling method were able to enhance the sampling of phase space in RNA tetraloop simulations.

We investigated whether the combined GB + LES approach was able to provide stable simulations of the correct structure under conditions that also resulted in spontaneous conversion of incorrect to correct structure. In addition, we investigated whether any advantage is gained by using GB + LES as compared to GB alone. In other words, are the enhancements provided by GB and LES complementary?

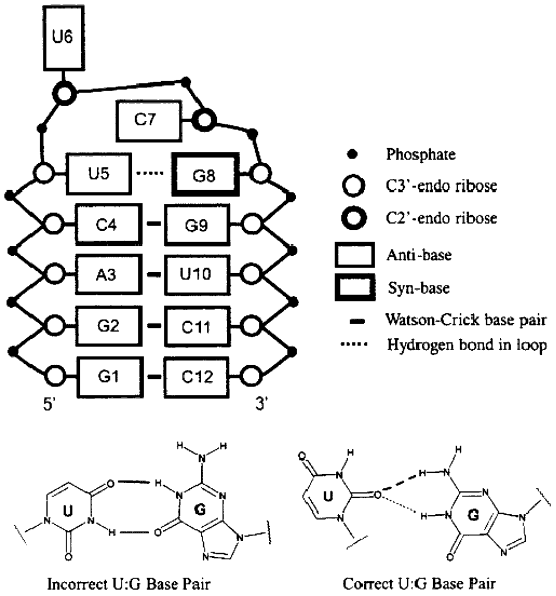

Figure 3. A schematic diagram of the topology of the RNA tetraloop system being studied. The hydrogen bond patterns for the U5:G8 base pair in the incorrect and correct NMR structures are shown in the lower figure. Solid lines are used for the incorrect hydrogen bonds, and dashed lines are used for the correct ones.

TABLE 2: Summary of Results Obtained from Various Simulations for the RNA Tetraloop[a]

	no. LES copies	temp (K)	total no. of atoms	starting structure	time of I→C transition (ps)	final RMSD to C
PME + LES[b]	5	300	7358	I	200	0.8
GB non-LESa	1	300	382	C	n/a	0.9
GB non-LESb	1	300	382	I	1100	1.0
GB non-LESc	1	300	382	I	1600	1.0
GB non-LESd	1	300	382	I	200	1.0
GB + LESa	3	200	632	C	n/a	1.0
GB + LESb	3	200	632	I	40−370	1.0
GB + LESc	3	200	632	I	160	1.0
GB + LESd	3	130	632	I	470−1340	1.0
GB + LESe	3	100	632	I	n/a	3.0

[a] The details of each simulation are described in the main text. Transition times for LES simulations are given as a range when the different copies showed significantly different transition times. [b] Reference 32.

The results of all simulations with various temperature and solvent models from both I and C structures are summarized in Table 2. All root-mean-square deviation (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 RMSD selection was chosen to be consistent with the LES/PME study of this system.[32] The results of each simulation are described in further detail below.

We first made the same choice as Simmerling et al. did in the previous LES/PME study of this RNA tetraloop and used five LES copies of the entire UUCG loop.[32] Each of these copies was attached to the stem, and the stem interacted with these copies in an average way. We initiated a simulation at 300 K from the correct structure, with all LES copies having identical initial coordinates. This simulation resulted in fully extended conformation of RNA within 400 ps. The heavy atom RMSD rose to 4.5 Å and all base pairs were lost during the simulation. Similarly undesirable results were obtained when starting from the incorrect conformation. We hypothesized that this might

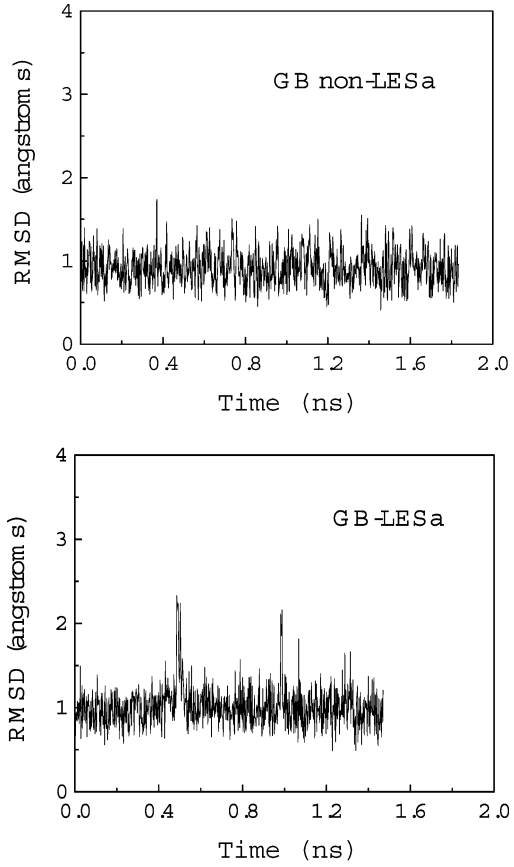

Figure 4. The loop RMSD as a function of time for simulations starting with the correct conformation. On the top is the GB simulation (GB non-LESa in Table 2); on the bottom is the GB + LES simulation (GB−LESa in Table 2). Only average LES RMSD is displayed for clarity.

arise from too great a weakening of the Watson−Crick hydrogen bonds because of the scaling of partial charges and Lennard-Jones well depth parameters. It has also been shown that the behavior of the LES system corresponds to a non-LES system of higher temperature.[44] Moreover, lack of solvent friction likely makes the dynamic behavior of the RNA more sensitive than with LES in explicit solvent.

We therefore empirically reduced the temperature and number of copies that were used and evaluated the dynamics of the correct structure. We found that three LES copies at 200 K (GB + LESa in Table 2 and lower graph in Figure 4) resulted in a stable simulation with similar fluctuations to those observed in non-LES GB simulation at 300 K (GB non-LESa and upper graph in Figure 4). In both cases, the structure was stable and the loop RMSD values fluctuated about 1 Å. We also obtained stable trajectories of the correct conformation with three LES copies at temperatures of 150 K and 130 K (data not shown). The use of LES has therefore not affected the ability of the simulation to maintain a stable correct conformation (although with a reduced temperature).

This scaling of temperature makes the use of GB + LES (and LES in general, although GB + LES appears to be particularly sensitive) somewhat more complex than standard MD. Others have pointed out this difficulty with LES, and we do not address this aspect of the LES method here; rather, we are interested in combination of the method with an efficient solvation model. As a general guide to evaluating temperature when predicting unknown structure, we expect that the LES copies should converge to a single conformation regardless of initial structure. When optimization protocols such as simulated

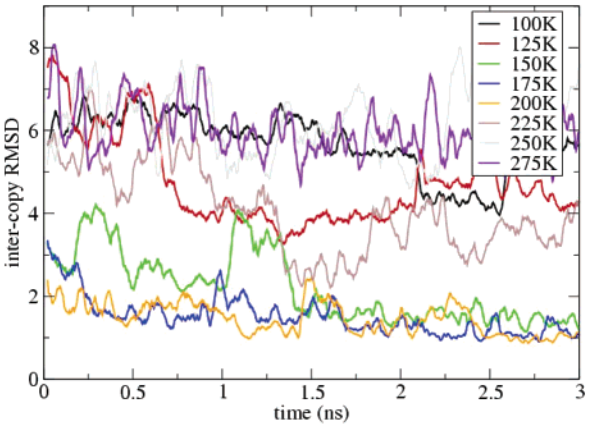

Figure 5. Average all-atom deviations between the three LES copies during GB + LES simulation at various temperatures. All simulations were initiated with the same set of alternate LES conformations and the initial deviation between the copies was ~4.5 Å. All of the copies converge to a single family of structures in 1−1.5 ns at 150, 175, and 200 K.

annealing are used, multiple predictions should be compared to ensure that the final structures are not the result of kinetic trapping. One should always demonstrate insensitivity of results to the initial coordinates for well-converged simulations of any type. With LES, we can compare the predictions being provided by each LES copy as the simulation proceeds. This provides us with continuous measures of the precision of the prediction by exploiting the built-in convergence test available with LES, especially when using alternate initial conformations for the copies. One can thus obtain conformational "error bars" in a single LES simulation.

To demonstrate this approach, we initiated simulations with three LES copies of the loop region and assigned each copy a different initial structure chosen at random from high-temperature dynamics. We calculated the average all-atom deviation of the copies *from each other* during the simulation (fit to the noncopied stem), and present the average RMSD of the copies as a function of time in Figure 5. This procedure was repeated for temperatures in the range of 100−275 K in increments of 25 K. This gives us a temperature-dependent measure of how quickly the copies converge to a similar set of structures, regardless of what that structure may be. At low temperatures, such as 100 K, the deviation remains high because the barriers are still too large to overcome on this time scale even with LES. At 150, 175, and 200 K, all of the copies converge to a set of structures that differ from each other by <2 Å for all atoms, even though they started from different conformations. At higher temperatures, the deviation remains large not because of trapping, instead the copies show large motions but do not sample a single structure. This suggests that a useful range for obtaining a converged prediction is 150−200 K.

Without LES, one must repeat the simulations from different initial conditions for each optimization protocol to get a measure of its reliability, although we can and do carry out independent LES simulations to gain further confidence in their convergence as well. We are currently investigating alternate approaches to aid in selecting the optimal temperature or the use of replica exchange approaches[45] to avoid the need to select a single temperature.

Next, simulations were carried out starting from the incorrect conformation under conditions in which the correct conformation was stable. In the non-LES GB simulation (GB non-LESb, upper graph in Figure 6), the RNA underwent conformational change at about 1100 ps, a time scale very similar to the 1200 ps

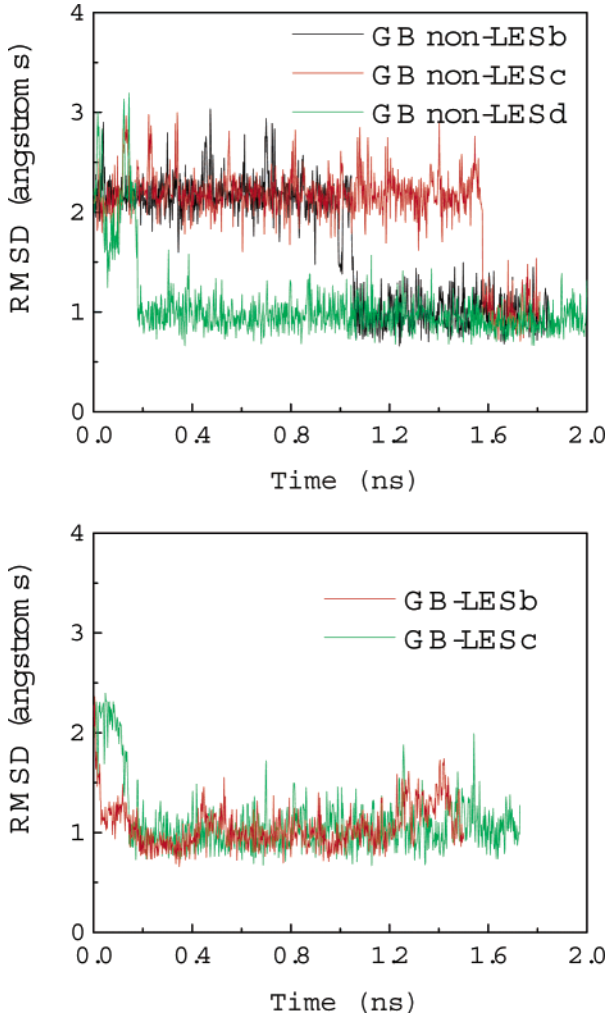

Figure 6. The loop RMSD (compared to the correct structure) as a function of time. On the top are three independent non-LES GB simulations (GB non-LESb, c, d). On the bottom are two GB + LES simulations (GB−LESb and GB−LESc). Three LES copies were employed for the entire UUCG loop. The transition from incorrect to correct structure occurs much more rapidly in the LES simulations.

reported for non-LES GB by Williams and Hall. The hydrogen bond geometry started to reorganize in less than 400 ps, but the successful transition was not achieved until about 1 ns (as measured by RMSD). The critical conformational change involved rotation of the U5 N1−C1′ torsion angle and took place over a short time scale (<10 ps, as indicated by the sudden drop of RMSD from 2.2 Å to about 1 Å).

To gain a better understanding of the pathways of this transition, we performed three additional simulations that differed only in the assignment of initial velocities. Analysis of the trajectory data indicates that two of these simulations (GB non-LESc and d) showed similar transition pathways (but different time scales, shown in the upper graph of Figure 6). However, the other non-LES simulation failed to convert to the correct structure even after 10 ns (data not shown). During that simulation, a large change in backbone conformation was observed near C7, and the U5 base partially flipped out of the loop.

In the lower graph of Figure 6, results from two alternate GB + LES simulations at 200 K are shown (GB + LESb and c). Similar to that observed for non-LES GB, a reduction in RMSD from 2.3 to 1.0 Å occurs in both simulations, demonstrating that these GB + LES simulations achieved the transition

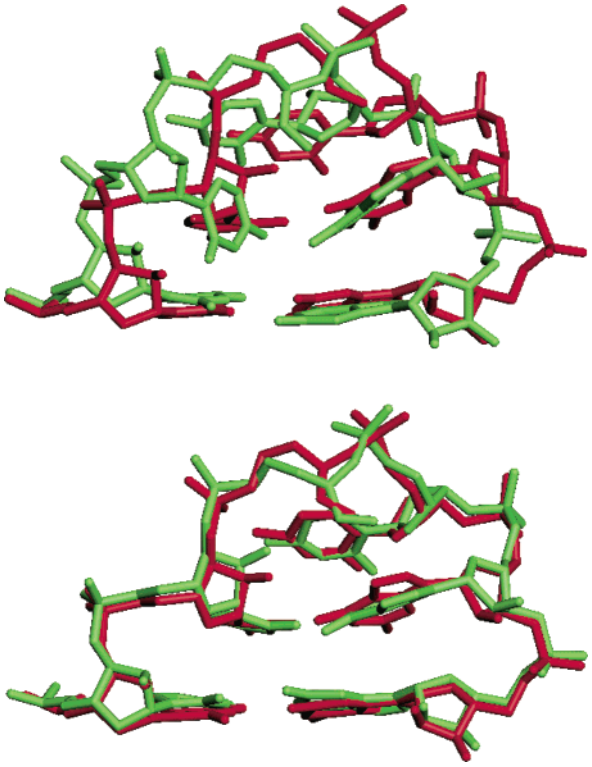

Figure 7. Comparison of the loop regions in the correct NMR structure (red) and the average LES MD structure (green) from GB simulations using three LES copies of the UUCG loop region. Only the loop and the first base pair of the stem are shown (residues 4−9 except the mobile U6 base). The upper image shows the initial (incorrect) structure. In addition to the base pair hydrogen bond differences shown in Figure 3, 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 loop. The lower image shows the same comparison after 500 ps of GB + LES molecular dynamics simulation (GB−LESb). All of the major differences have been corrected.

from incorrect to correct conformation. However, the transition occurs on a substantially shorter time scale with LES than for the single-copy simulations. These RMSD values represent an average value for the entire three-copy LES system; each of the three copies converted to the correct NMR structure, and the details of the transition for each copy will be discussed in further detail below. Figure 7 shows the comparison of the loop regions in the correct NMR structure and the average LES MD structure from GB + LES simulation after 500 ps of molecular dynamics simulation. All of the major differences from the initial (incorrect) structure have been corrected. Similar to observations based on PME + LES and standard GB simulations, the U6 base samples multiple conformations even when the remainder of the stem-loop system samples the correct geometry.

We further examined the sensitivity of these results to the simulation temperature. We expected and observed that at a lower temperature (130 K, GB + LESd) the transition takes place on a longer time scale than at 200 K. However, when the temperature was reduced to 100 K, the C conformation remained stable for the LES copies but the I conformation was also stable, failing to convert to the C conformation during the 4 ns simulation (GB−LESe). This is consistent with our previous observation that the alternate LES conformations did not converge to a single structure at this temperature (Figure 5). Thus, the structure was kinetically trapped under these conditions even with LES.

One must be cautious in the use of the time of a single observed transition to represent the actual rate that would be

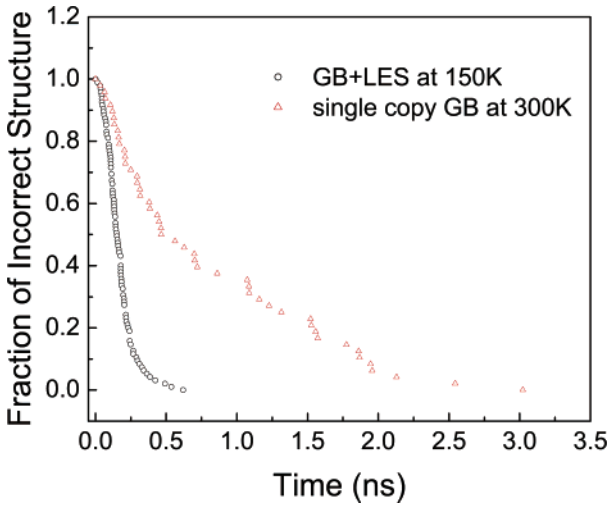

Figure 8. RNA tetraloop folding kinetics characterizing the conversion of an ensemble of incorrect structures to the correct one. The red triangles denote 48 independent single-copy GB simulations at 300 K. The black circles represent 96 GB + LES trajectories at 150 K.

obtained for an ensemble of such events. Likewise, comparison of times obtained from relatively few events using different methodologies may not reflect the actual differences in barrier crossing rates. We therefore performed more statistically significant comparisons of the rate of I→C transition from standard GB and GB + LES simulations.

A series of simulations were carried out for the incorrect structure, differing only in initial velocity assignments. For standard GB, 48 independent trajectories were obtained at 300 K. Similarly, 32 GB + LES simulations were performed at 150 K, each using three copies, thus providing 96 loop trajectories. In each case, the time dependence of first passage (I→C) was collected until the *entire* ensemble had undergone the I→C transition (Figure 8). This ensures that differences in rates are due solely to reduction in free-energy barriers and are not the result of variance in transition times because of poor statistical sampling. The process is clearly more rapid in GB + LES than for standard GB, with estimated rate constants of 6.5 ns^{-1} and 1.2 ns^{-1}, respectively.

As an additional test of convergence, we calculated the force metrics (eq 5) for the nonbonded forces acting on the tetraloop atoms as a function of time.[38−41] These values should approach zero for pairs of trajectories that are sampling the same regions of phase space. In the standard simulations (Figure 9), we see subnanosecond convergence for pairs of trajectories that were initiated from the same structure (both incorrect or both correct). However, convergence is not obtained for pairs of simulations started from two different structures. This indicates that these non-LES simulations remain confined to their initial conformational substates and are unable to cross the intervening barriers on this time scale.

A remarkable difference can be seen from the force metrics for the GB + LES simulations. In Figure 10, all of the nonbonded force metrics decay to 0 after ∼500 ps, irrespective of the initial conformations. These results provide additional evidence that GB + LES can explore different conformational substates on the nanosecond time scale more efficiently than standard GB simulations.

One advantage of LES is the ability to accelerate conformational transitions. Another potential advantage is that the copies may explore alternate regions in phase space, thus providing multiple transition events in a single simulation at a reduced computational cost as compared to multiple non-LES simula-

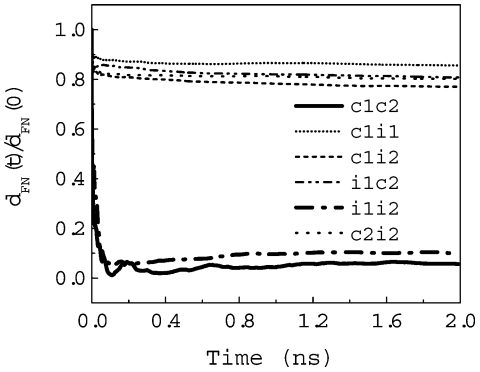

Figure 9. Plots of the normalized nonbonded force metric, $d_{FN}(t)/d_{FN}(0)$, as a function of time for the RNA tetraloop during standard GB simulations at 300 K. Data are calculated for pairs of trajectories, denoted in the legend. c1 and c2 started from the correct structure; i1 and i2 started from the incorrect structure. The self-averaging metric approaches zero on this time scale only for trajectories initiated in the same structure.

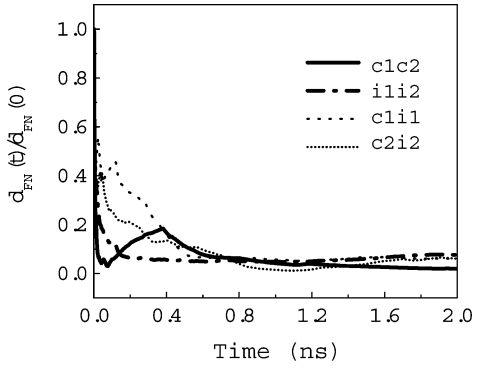

Figure 10. Plots of the normalized nonbonded force metric, $d_{FN}(t)/d_{FN}(0)$, as a function of time for the RNA tetraloop during GB + LES simulations at 150 K. Data are calculated for pairs of trajectories, denoted in the legend. c1 and c2 started from the correct structure; i1 and i2 started from the incorrect structure. In contrast to non-LES simulations, the metric approaches zero on this time scale for all pairs of trajectories.

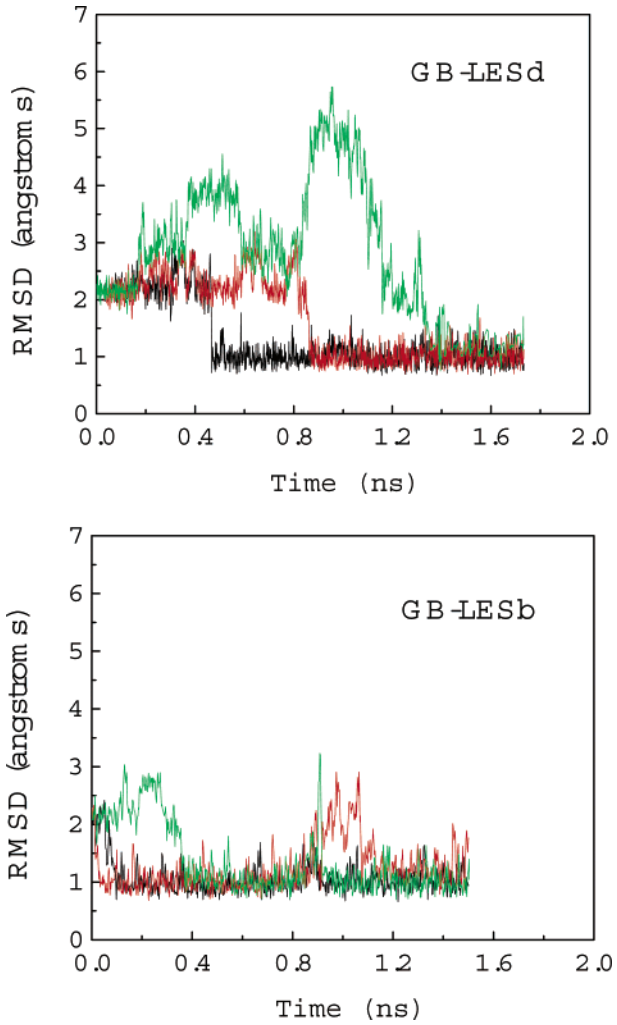

Figure 11. The loop RMSD (compared to the correct structure) as a function of time for each of the three LES copies. Both simulations were initiated from the incorrect structure. The copies undergo the transition to the correct structure at different times. The same behavior is seen during simulation at 130 K (upper) and 200 K (lower).

tions. In our previous investigation of this I→C transition in explicit solvent,[32] the instantaneous backbone RMSD values between pairs of different copies were less than 0.2 Å throughout the entire simulation. This indicates that the copies not only were unable to explore alternate transition pathways, but a time coupling was also present and therefore only a single time scale for the event could be sampled.

As described above, our goal during our development of the GB + LES model was to overcome this weakness of LES through individual solvation of each copy, avoiding the caging effect associated with a single explicit solvent cavity for all copies. This directly affects the ability of the LES simulation to model alternate transition pathways in a single simulation. We therefore investigated whether our approach was successful in overcoming it. In Figure 11, we show the RMSD value for *each* of the copies as a function of time (in contrast to Figure 6 in which the average RMSD for all copies was shown) for GB + LES simulation at 130 K (GB−LESd) and 200 K (GB−LESb). It is apparent that each copy undergoes the I→C transition. In contrast to the simulation with explicit solvent, however, the copies show much greater independence and undergo the transition at significantly different times. Similar results were described above in which the deviations in copy conformations were large (Figure 5), but in that case the divergence was a product of the initial coordinate generation and in the present case the copies are able to spontaneously explore alternate pathways after starting from the same conformation.

The difference in peak RMSD values for each copy suggests that different transition pathways are explored in this single simulation. We examined the events for each copy in simulation GB + LESd in detail and show representative snapshots in Figure 12. In the first pathway (left column in Figure 12), the most direct interconversion is observed. At 100 ps, the U5−G8 base pair partially lost the original hydrogen bond pattern because of the rotation of the G8 base pair about N9−C′ dihedral. U5 is then observed to flip back and forth via rotation of N1−C1′ dihedral. At 400 ps, this flipping motion results in formation of partially correct structure. Meanwhile, 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 also are achieved, as shown by the hydrogen bond plot (Figure 13). At this point, the hydrogen bond pattern is correct while the relative orientation of these bases and the stacking of the UG pair against the stem CG pair differ from that found in the correct conformation. At 470 ps, both the correct hydrogen bond pattern and stacking are attained. This conformation is retained throughout the remainder of the simulation.

The middle column of Figure 12 shows a similar transition pattern for the second copy. However, the reorganization of the

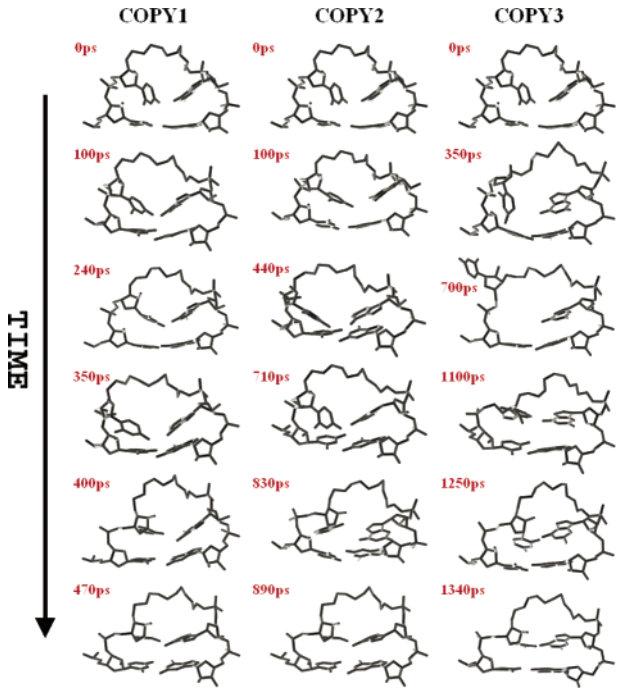

Figure 12. Snapshots of the U5:G8 base pair during simulation GB + LESd that employed three LES copies of the UUCG loop. Also shown is the C4:G9 base pair in the stem, which has a stacking interaction with the U5:G8 pair. The three columns correspond to snapshots of each of the three sets of copies. For clarity, the rest of the system is not drawn. The details of the transitions are provided in the main text.

two U5:G8 bases after partial separation takes ~600 ps, and the base pair reformed with the correct hydrogen bond pattern but with stacking against the stem that is similar to the incorrect conformation. The correct stacking pattern and relative orientation were attained within the next 60 ps. In this case, the most significant structural transition was also achieved by rapid rotation of base U5 N1−C1′ dihedral angle, but the backbone near C4 and U5 shows significantly greater distortion during the transition as compared to the first pathway.

In contrast, the right column shows a dramatically different interconversion sampled by the third copy. At 100 ps, the U5−G8 bases separate and fluctuate back and forth for about 600 ps. At 700 ps, the U5 base flips out of the loop, remaining in this location for more than 400 ps, finally shifting back toward the loop through a rotation of the C3′−C4′ dihedral. The base is then positioned to form the correct hydrogen bonds and converts to the correct structure through shifts in stacking that occur over the next 100 ps.

Detailed comparison of the pathways sampled by the LES copies with those sampled during non-LES GB simulations revealed that the first two LES pathways are both remarkably similar to those observed in all three successful GB non-LES simulations, despite the difference in the time scale of the transitions. The third LES pathway shown in Figure 12 was not observed in the non-LES GB simulations. However, a flipping motion of the U5 base similar to that which initiated this transition was observed in the non-LES GB simulation described above that did not convert to the correct structure. A similar flipping motion of the U5 base was observed in standard MD simulation of this system in explicit solvent (Miller, J., pers. comm.), suggesting that the process may be involved in a pathway of lower probability. Further investigation is in progress.

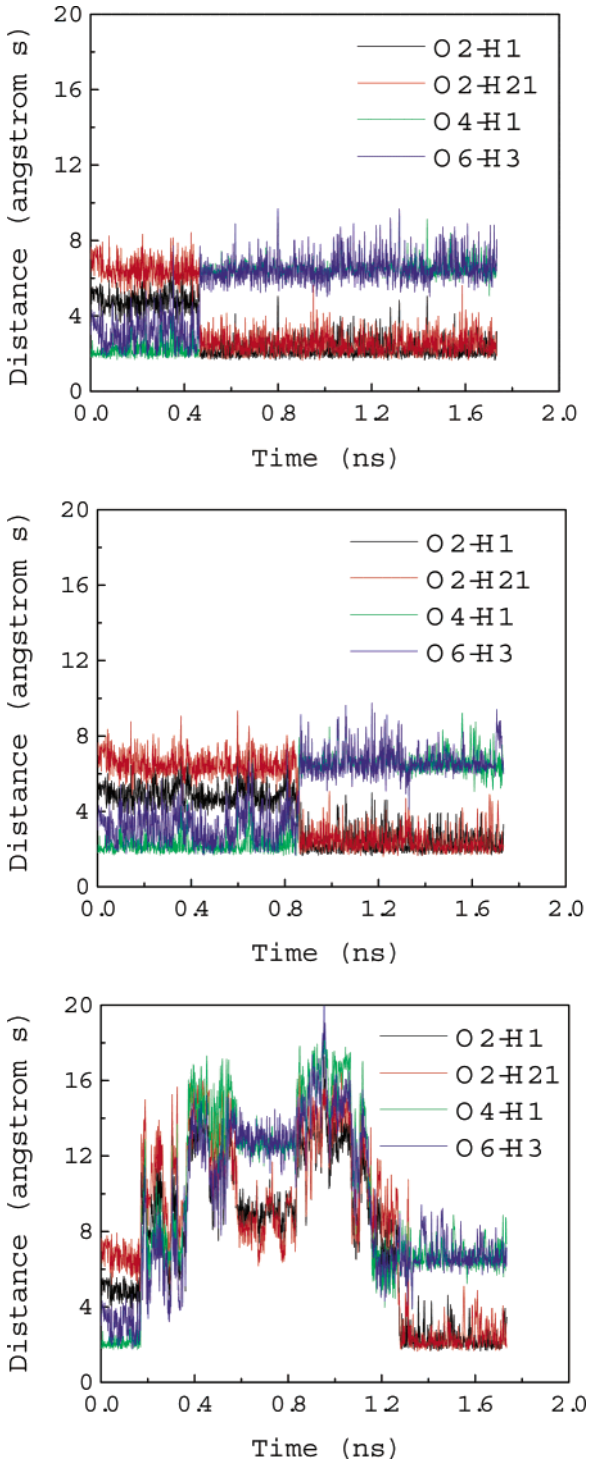

Figure 13. Distances between hydrogen and oxygen atoms corresponding to the hydrogen bonds in the U5:G8 base pair shown in Figure 3. Green and blue are distances of the initial incorrect hydrogen patterns; red and black are distances of the final correct hydrogen patterns. The three figures represent the three loop copies and correspond to the three columns in Figure 12.

That these different transitions occur as three independent events observed in a single LES simulation is quite remarkable and further demonstrates the advantages of this combined methodology: most current simulation methods do not allow for even a single observation of such events and therefore cannot provide rapid insight concerning the existence and nature of alternate transition pathways.

Conclusions

We have shown for the first time that the GB model can be used to provide a proper implicit solvation treatment for a simulation employing locally enhanced sampling. We derived an exact approach to combining the two algorithms in which each LES copy is individually solvated, rather than the solvation of the ensemble of copies that is obtained with explicit solvation. Our approach does lead to increased calculations in the non-LES region (not seen with PME LES) but a reasonable approximation reduced this overhead by ~70%. This approach has been implemented in the AMBER suite of programs and will be available in the next AMBER release (version 8).

We carried out tests of the combined approach using a well-studied RNA UUCG tetraloop system. Previous work had shown that the use of either GB solvation or LES in explicit solvent was able to successfully model the conversion of an incorrect to correct conformation. We have shown that the GB + LES simulations have the same ability to reproduce the correct conformational change, but with greater computational efficiency than obtained with either GB or LES alone. This improved convergence was demonstrated by (1) comparison of rate constants for the conversion obtained from a large set of independent simulations and (2) calculation of force metrics that clearly show improved sampling in the LES simulations.

The reduction in barrier heights provided by LES results in simulations that are much more efficient than single-copy GB simulations. However, these combined GB + LES simulations are more sensitive to the temperature and number of copies than corresponding non-LES GB simulations or LES simulations carried out in the explicit solvent. We showed that one way to evaluate these parameters for a system in which the correct structure is unknown is to monitor the convergence of the conformations of the alternate copies after assigning them different initial coordinates.

Finally, we make the remarkable observation that a single LES simulation can provide multiple (and qualitatively different) instances of key transitions. LES simulations in explicit solvent for this system were unable to provide independent transitions for the copies because of a caging effect arising from the sharing of a single solvent cavity for all LES copies. We believe that this approach is likely to be an important component of all-atom structure refinement of biomolecular systems, particularly when a portion of the structure, such as a loop region, is poorly determined and requires additional local sampling.

Acknowledgment. C.S. thanks Adrian Roitberg for useful discussions. Support for this project was provided by the AMDeC Foundation of New York City, through its "Tartikoff/Perelman/EIF fund for Young Investigators in Women's Cancers", and by NIH GM6167803. Portions of this work were carried out at the Pittsburgh Supercomputing Center through NPACI grant MCA02N028 to C.S.

References and Notes

(1) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(2) Jorgensen, W. L.; Maxwell, D. S.; Tiradorives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(3) 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.

(4) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(5) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.

(6) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(7) *Interaction models for water in relation to protein hydration*; Berendsen, H. J. C., Vangunsteren, W. F., Postma, J., Hermans, J., Eds.; Reidel: Dordrecht, The Netherlands, 1981; p 331.

(8) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

(9) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(10) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(11) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.*; **2001**, *98*, 14931.

(12) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345.

(13) Brooks, C. L. *Acc. Chem. Res.* **2002**, *35*, 447.

(14) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(15) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(16) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243.

(17) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.

(18) Williams, D. J.; Hall, K. B. *Biophys. J.* **1999**, *76*, 3192.

(19) Miller, J. L.; Kollman, P. A. *J. Mol. Biol.* **1997**, *270*, 436.

(20) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.

(21) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161.

(22) Czerminski, R.; Elber, R. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 70.

(23) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277.

(24) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534.

(25) Simmerling, C. L.; Elber, R. Proc. *Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3190.

(26) Stultz, C. M.; Karplus, M. *J. Chem. Phys.* **1998**, *109*, 8809.

(27) Hornak, V.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 577.

(28) Case, D. A.; Pearlman, D. A.; Caldwell, J. A.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 6*; University of California: San Francisco, CA, 1999.

(29) Cui, G.; Simmerling, C. *J. Am. Chem. Soc.* **2002**, *124*, 12154.

(30) Varani, G.; Cheong, G.; Tinoco, I. J. *Biochemistry* **1991**, *30*, 3280.

(31) Allain, F. H. T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333.

(32) Simmerling, C.; Miller, J. L.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 7149.

(33) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277.

(34) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426.

(35) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.

(36) Guvench, O.; Shenkin, P.; Kolossvary, I.; Still, W. C. *J. Comput. Chem.* **2002**, *23*, 214.

(37) Onufriev, A.; Case, D.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.

(38) Thirumalai, D.; Mountain, R. D.; Kirkpatrick, T. R. *Phys. Rev. A* **1989**, *39*, 3563.

(39) Thirumalai, D.; Mountain, R. D. *Phys. Rev. A* **1990**, *42*, 4574.

(40) Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 809.

(41) Straub, J. E.; Thirumalai, D. *Proteins* **1993**, *15*, 360.

(42) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.

(43) Lolis, E.; Alber, T.; Davenport, R. C.; Rose, D.; Hartman, F. C.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6609.

(44) Straub, J. E.; Karplus, M. *J. Chem. Phys.* **1991**, *94*, 6737.

(45) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.