

Article

Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations

Ramu Anandakrishnan,¹ Aleksander Drozdetski,² Ross C. Walker,³ and Alexey V. Onufriev^{1,2,*}¹Department of Computer Science and ²Department of Physics, Virginia Tech, Blacksburg, Virginia; and ³San Diego Supercomputer Center and Department of Chemistry and Biochemistry, University of California, San Diego, San Diego, California

ABSTRACT Adequate sampling of conformation space remains challenging in atomistic simulations, especially if the solvent is treated explicitly. Implicit-solvent simulations can speed up conformational sampling significantly. We compare the speed of conformational sampling between two commonly used methods of each class: the explicit-solvent particle mesh Ewald (PME) with TIP3P water model and a popular generalized Born (GB) implicit-solvent model, as implemented in the AMBER package. We systematically investigate small (dihedral angle flips in a protein), large (nucleosome tail collapse and DNA unwrapping), and mixed (folding of a miniprotein) conformational changes, with nominal simulation times ranging from nanoseconds to microseconds depending on system size. The speedups in conformational sampling for GB relative to PME simulations, are highly system- and problem-dependent. Where the simulation temperatures for PME and GB are the same, the corresponding speedups are approximately onefold (small conformational changes), between ~1- and ~100-fold (large changes), and approximately sevenfold (mixed case). The effects of temperature on speedup and free-energy landscapes, which may differ substantially between the solvent models, are discussed in detail for the case of miniprotein folding. In addition to speeding up conformational sampling, due to algorithmic differences, the implicit solvent model can be computationally faster for small systems or slower for large systems, depending on the number of solute and solvent atoms. For the conformational changes considered here, the combined speedups are approximately twofold, ~1- to 60-fold, and ~50-fold, respectively, in the low solvent viscosity regime afforded by the implicit solvent. For all the systems studied, 1) conformational sampling speedup increases as Langevin collision frequency (effective viscosity) decreases; and 2) conformational sampling speedup is mainly due to reduction in solvent viscosity rather than possible differences in free-energy landscapes between the solvent models.

INTRODUCTION

Molecular dynamics (MD) simulations are routinely used to study the structure, function, and activity of biological molecules (1–4). Over 12,000 articles regarding MD were published in 2009 alone, with over 300 in the top scientific journals (5). However, without highly specialized supercomputers (6), simulation times accessible by the most commonly used atomistic methods—those that provide the highest level of detail—are still much shorter than the time-scale of many important biomolecular processes, such as ligand binding, folding of most proteins, and enzyme turnover, which occur on timescales in the range of tens of microseconds to seconds and even longer (2,7–9). Without sufficiently long simulation times, these methods will most likely fail to sample some important conformations and structural transitions.

Atomistic MD simulation methods can be divided into two broad classes: those that treat solvent explicitly and those that treat solvent implicitly (10,11). The main objective of this study is to compare the speed of conformational sampling within these two very different approaches to

treating solvent effects. We also investigate the purely computational speedup due to the algorithmic differences between two commonly used explicit and implicit solvent models—the particle mesh Ewald (PME) explicit solvent model and the generalized Born (GB) implicit solvent model.

Explicit-solvent methods, without further approximations, treat solvent molecules explicitly, i.e., interactions between all pairs of solute and solvent atoms are explicitly computed. The PME approximation, the most commonly used explicit-solvent method for biomolecular simulations, speeds up these computations by imposing an artificial periodicity on the entire system and treating the system as an infinite crystal with identical repeating cells (12–15). This assumption allows for a mathematical transformation that approximates the long-range interactions very efficiently, without significant loss in accuracy.

Implicit-solvent methods, on the other hand, speed up atomistic simulations by approximating the discrete solvent as a continuum, thus drastically reducing the number of particles to keep track of in the system. An additional effective speedup often comes from much faster sampling of the conformational space afforded by these methods. The GB approximation, the most commonly used implicit-solvent method in atomistic MD simulations, approximates

Submitted April 16, 2014, and accepted for publication December 29, 2014.

*Correspondence: alexey@cs.vt.edu

Editor: Nathan Baker.

© 2015 by the Biophysical Society
0006-3495/15/03/1153/12 \$2.00



long-range electrostatic interactions via an analytical formula (16–43):

$$E_{ij}^{\text{elec}} = E_{ij}^{\text{vac}} + E_{ij}^{\text{solv}} \quad (1)$$

$$E_{ij}^{\text{vac}} = \frac{q_i q_j}{r_{ij}} \quad (i < j) \quad (2)$$

$$E_{ij}^{\text{solv}} = -\frac{1}{2} \left[\frac{1}{\epsilon_{\text{in}}} - \frac{\exp(-0.73\kappa f_{ij}^{\text{GB}})}{\epsilon_{\text{out}}} \right] \frac{q_i q_j}{f_{ij}^{\text{GB}}} \quad (3)$$

$$f_{ij}^{\text{GB}} = \sqrt{r_{ij}^2 + B_i B_j e^{(-r_{ij}^2/4B_i B_j)}}, \quad (4)$$

where E_{ij}^{elec} , E_{ij}^{vac} , and E_{ij}^{solv} are the total, vacuum, and solvation energy contributions due to electrostatic interactions between atoms i and j . Here, ϵ_{in} and ϵ_{out} are the internal (solute) and external (solvent) dielectric constants, q_i and q_j are the charges of atoms i and j , r_{ij} is the distance between the atoms, and B_i and B_j are their effective Born radii. The effective Born radii account for the dielectric screening effect of the solvent, and their values reflect the atom's degree of burial within the solvent. The $\exp(-0.73\kappa f_{ij}^{\text{GB}})$ term accounts for the screening effect of monovalent salt (44), with κ being the Debye-Hückel screening parameter.

Implicit treatment of solvent, although arguably less accurate than the explicit-solvent approach (45,46), is widely used in molecular simulations for two main reasons. First, the implicit-solvent method can be algorithmically/computationally faster, as measured by simulation time steps per processor (CPU) time, because the vast number of individual interactions between the atoms of individual solvent molecules do not need to be explicitly computed. Second, and now perhaps the main reason for the use of implicit- instead of explicit-solvent simulations, is that implicit-solvent simulations can sample conformational space faster (25,47–49). To some extent, the interest in implicit-solvent-based simulations is motivated by the need to sample very large conformational spaces for problems such as protein folding (50,51), binding-affinity calculations (48), or large-scale fluctuations of nucleosomal DNA fragments (52).

Two main factors are expected to contribute to the speedup of conformational sampling by implicit-solvent models such as the GB, compared to explicit-solvent models: 1) reduction of the effective solvent viscosity, which, ideally, has no effect on the energy landscape (thermodynamics) of the system; and 2) possible alterations of the energy landscape itself, some of which may have desirable speedup effects on kinetics but undesirable effects on thermodynamics (53,54). These issues will be discussed in the Results and Discussion section below, with the most detailed analysis presented for the case of miniprotein

folding. This study does not investigate the accuracy of the implicit-solvent approximations, explicit water models, PME, force fields, etc. These very important issues have been extensively studied elsewhere (34,55–60). Comparing the accuracy of MD methods on realistic bimolecular structures is a complex, multidimensional problem; there is little hope of arriving at a full picture in a single study. However, the dimensionality of the problem is greatly reduced for efficiency (speedup) comparisons, so we are confident that careful tests on a few judiciously chosen molecular systems will give us a reasonably comprehensive picture, at least for the two commonly used solvent models we are going to analyze.

Several previous studies (25,47,48,51) have demonstrated that the speed of conformational sampling in implicit-solvent simulations can be ~2–20 times faster than common explicit-solvent PME simulations. However, a number of important questions remain: 1) How does the speedup of conformational sampling depend on structure size and type of conformational change involved? The sampling speed is affected by an intricate balance between solute-solute and solute-solvent friction (61,62), and this balance is expected to be different for different types of systems and conformational transitions (47). 2) How do simulation parameter(s) that determine the effective solvent viscosity influence the speedup? The influence of solvent viscosity on kinetics has not been thoroughly investigated beyond small proteins (51,61,62) and peptides (47,63). 3) Is the speedup primarily due to reduced viscosity or to possible alterations of the free-energy landscapes by the implicit-solvent treatment? 4) What is the combined speedup—the rate of conformational sampling per day of real simulation time—when the algorithmic differences between realistic explicit- and implicit-model implementations are taken into account? After all, it is this effective speed that is of direct interest to practitioners. To address all of these questions, we have performed a systematic analysis across a range of structure sizes and types of conformational transitions.

The types of conformational changes considered here include dihedral angle flips in a protein, nucleosome tail collapse, unwrapping of DNA from the nucleosome core, and folding of a miniprotein. The solute structure sizes considered here range from 166 to 25,100 atoms.

The two speed-up methods we have chosen for comparison, the GB and the PME methods, are by no means unique. Many other approaches exist for increasing computational efficiency of atomistic MD simulations and for speeding up conformational sampling (for reviews, see, e.g., Onufriev (10) and Zuckerman (45)). Replica-exchange MD, for example (64–66), enhances conformational sampling by exchanging conformational samples across simulations at multiple temperatures. Accelerated MD simulations enhance conformational sampling by lowering the energy barrier between conformational states (46). The use of longer integration time steps for certain parts of the

calculation can increase the simulation time per CPU time (67–69). However, these and many other speedup and enhanced-sampling methods can be applied to both the explicit- and implicit-solvent models. Since the focus of this work is comparing the effects of the two generic opposing solvent models on the efficiency of biomolecular simulations, we do not consider these additional speedup and enhanced-sampling methods here.

The remainder of this article is organized as follows. The structure preparation and simulation protocol used for the simulations are described in the Materials and Methods section. The results of the simulations are discussed in the Results and Discussion section. And our findings are summarized in the Conclusions.

MATERIALS AND METHODS

We used five different molecular structures for this study: a 166-atom protein, CLN025 (70), a more stable version of chignolin (71), to study folding; a 4812-atom phospholipase C (Protein Data Bank (PDB) (72) ID 1GYM (73)) to study dihedral angle flips; a 25,100-atom nucleosome complex (PDB ID 1KX5 (74)) to study the collapse of histone tails onto the DNA and the unwrapping of the DNA from the histone core. The structure preparation and simulation protocols used for each of the above structures vary slightly from structure to structure. However, for the most part, the protocol described below was used, with variations from this protocol noted in the Methodological details section.

The protonation state for titratable groups were set using the H++ web server (75) for pK prediction and structure preparation, at pH 6.5. For explicit-solvent simulations, the structures were solvated in a truncated octahedral TIP3P (76) water box extending 10 Å from the solute. Counterions were added to the water box to neutralize the system while approximating a salt concentration of 0.145 M. The Amber ff10 (77) force field was used for both implicit- and explicit-solvent simulations.

To run the MD simulations, we used the GPU implementation of the Amber 12 MD software package with the SPFP model (78–80). The PME method (12) with constant-volume periodic boundary condition was used for explicit-solvent simulations and the OBC variant (igb5) of the GB method (16) for the implicit-solvent simulations. All references to PME in this article refer to the explicit-solvent PME method with a TIP3P water box. The simulation protocol consisted of five stages. First, the structure was relaxed with 2000 steps of conjugate-gradient energy minimization, with solute atoms restrained to the initial structure by a force constant of 5 kcal/mol/Å². Next, the system was heated to 300 K over 600 ps, with a restraint force constant of 1 kcal/mol/Å². The system was then equilibrated for 2 ns with a restraint force constant of 0.1 kcal/mol/Å², followed by another 2 ns with a restraint force constant of 0.01 kcal/mol/Å². All restraints were removed for the production stage. The simulation time step was 2 fs/iteration. A direct space cutoff of 8 Å was used for all stages of the PME simulations. No cutoff was used for the calculation of pairwise interactions, with $rgbmax = 15$ Å. Langevin dynamics (81) was used for temperature regulation with a collision frequency of $\gamma = 0.01$ ps⁻¹. The Shake algorithm (82) was used to constrain covalently bound hydrogen atoms. For the analyses presented in the Results and Discussion section below, snapshots of the MD trajectory were taken every 10 ps. Default values were used for all other simulation parameters. See the [Supporting Material](#) for additional details.

RESULTS AND DISCUSSION

Three different types of conformational transitions were considered here: small (dihedral angle flips), large

(nucleosome histone tail collapse and DNA unwrapping from the nucleosome histone core), and mixed (folding of a miniprotein). Our choice of representative test structures was dictated by the need to run simulations of sufficient length to adequately sample the relevant conformation space in each representative case.

The total conformational sampling speedup offered by the implicit-solvent model, compared to the explicit-solvent alternative, may be due, at least in part, to the altering of the free-energy landscape, which may affect the underlying thermodynamics of the problem. Whether or not it is appropriate to use implicit solvent to benefit from the associated speedups depends on the problem at hand: for each of the conformational transitions considered here, we note the type of problem for which the model may still be useful despite the altered free-energy landscape.

In the following, we report standard error (SE) along with the mean value as mean \pm SE. The computation of standard error is described in the [Supporting Material](#).

Small conformational changes

Dihedral angle flips

Many biological processes, such as nonallosteric protein-ligand binding, involve small local conformational changes. To study such small conformational changes, we analyzed variations in χ_1 and χ_2 dihedral angles in explicit-solvent (TIP3P) PME and implicit-solvent GB simulations of an ~300 residue protein, 1GYM.

For the 770 ns simulations considered here, the GB simulation does not explore exactly the same conformations as the explicit solvent (TIP3P) PME simulation. For example, the backbone root mean-square deviation (RMSD) relative to the starting structure is 1.6 Å for the explicit-solvent (TIP3P) PME simulation compared to 4.5 Å for the GB simulation (see [Supporting Material](#)). Therefore, to exclude, or at least reduce, the effect of larger global changes, we only considered groups where the distribution of the χ_1 and χ_2 angles were similar for the GB and explicit-solvent (TIP3P) PME simulations, i.e., the frequency at which the dihedral angle ranges were sampled differs by <10%. We identified 73 of 268 groups that met these criteria for χ_1 angles, and 53 of 213 groups for χ_2 angles.

To examine the sampling of χ_1 and χ_2 dihedral angles, we defined three ranges for all the dihedral angles (0–120°, 120–240°, and 240–360°) except the χ_2 angles for ASN, ASP, HIS, PHE, and TYR. For those angles, we used two ranges, 180–0° and 0 to +180°. The frequency of χ_1 (2.34 ± 0.002 flips/ns) and χ_2 flips (7.55 ± 0.006 flips/ns) for the GB simulation is on average only slightly faster than the χ_1 (1.58 ± 0.003 flips/ns) and χ_2 flips (5.15 ± 0.005 flips/ns) for the explicit-solvent (TIP3P) PME simulation ([Fig. 1](#)).

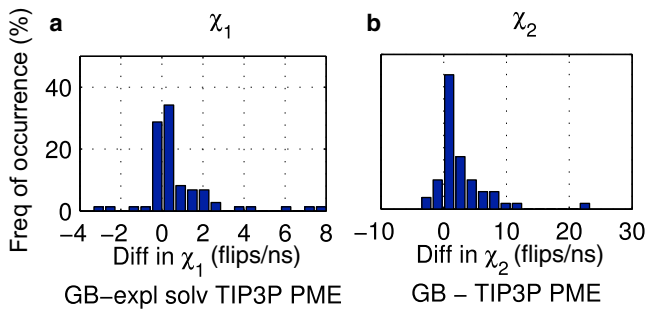


FIGURE 1 Difference in (a) χ_1 and (b) χ_2 dihedral angle flips (GB – explicit-solvent (TIP3P) PME) from 770 ns simulations of 1GYM protein. Flips are measured as changes in dihedral angles across distinct ranges of dihedral angles, as described in the text. To see this figure in color, go online.

Two conclusions can be made from examining Fig. 1: 1) on average, χ_1 and χ_2 flips occur only slightly (~1.5 times) more frequently with the GB simulation than with the explicit-solvent (TIP3P) PME simulation; and 2) in some cases, the reverse is true.

We considered several different factors that may explain why one method samples these dihedral angles more frequently than the other. We analyzed the relationship between frequency of dihedral angle flips and 1) depth of burial within the protein, as measured by the effective Born radii of the $C\beta$ and $C\gamma$ atoms for χ_1 and χ_2 angles, respectively (for an explanation, see the [Supporting Material](#)); 2) extent of hydrogen bonding between side chains and water, as measured by percentage of samples with such hydrogen bonds; 3) difference in side-chain conformation, as measured by residue RMSD; and 4) type of side chain. However, the correlations in each of these cases were not statistically significant (see [Supporting Material](#)), i.e., none of these factors could, separately, explain the differences in frequency of χ_1 and χ_2 dihedral angle flips. We speculate that some combination of these factors contributes to the differences in frequency of dihedral angle flips. For example, one would expect that the speedup for solvent-exposed surface groups would be higher than for buried groups. However, hydrogen bonds between side chains or steric constraints on large side chains, such as Tryptophan, can restrict the dihedral angles sampled, resulting in little or no difference in sampling speed between the implicit- and explicit-solvent models, even when the groups are on the surface.

Taking all of the above into account, we suggest that explicit solvent should be the primary choice for sampling dihedral transitions.

Large conformational changes

We considered two conformational changes to the 25,100 atom nucleosome, histone tail collapse and DNA unwrapping.

Nucleosome histone tail collapse

The nucleosome (PDB ID 1KX5) consists of an eight-protein histone core (H3, H3', H4, H4', H2A, H2A', H2B, and H2B') and a 147 bp DNA chain wrapped around the core (Fig. 2 a). In the PDB structure for 1KX5, the unstructured N-terminus tails for three of the histones (H3, H3', and H2B') extend well beyond the DNA surface. Experimentally, it has been shown that under physiological conditions, the unstructured positively charged histone tails collapse onto the negatively charged DNA (83). It has also been shown that histone tails regulate nucleosome mobility and stability (84), and MD simulations have been used to study this regulatory mechanism (85).

Both the explicit-solvent (TIP3P) PME and GB simulations reproduce the tail collapse (Fig. 2 b). The collapsed tail is highly unstructured and shows large conformational variation. For example, the number of histone H3 tail-DNA contacts fluctuates between 6 and 30 over the course of the simulation, even after the tail has collapsed. Since the question we are asking regards the general location of these highly unstructured tails, and not their specific conformations (which may also be highly dependent on the underlying gas-phase force field), the distance from the N-terminus of the histone to the geometric center of the DNA was used as the measure of histone tail extension.

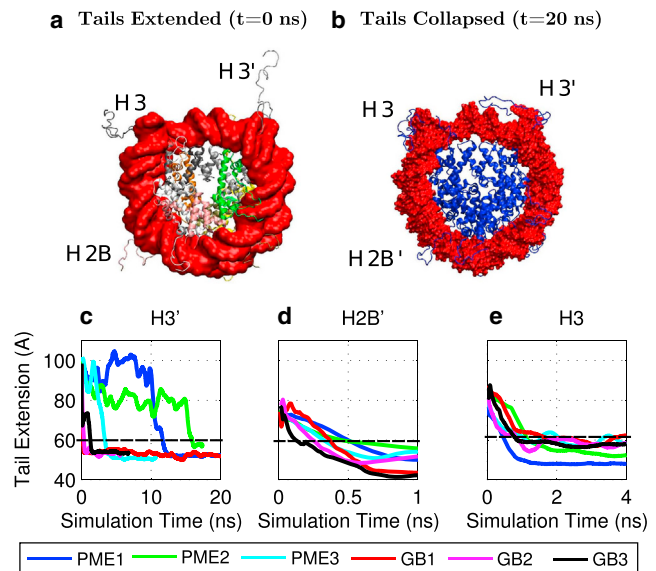


FIGURE 2 Three sets of independent MD simulations of the nucleosome tail collapse. (a) PDB structure with three histone tails (H3, H3', and H2B') extended. (b) Structure with the positively charged histone tails collapsed onto the negatively charged DNA. DNA surface is shown in red and the histone backbone structure is shown in blue. Tail extension is measured as the distance from the N-terminus tail to the center of geometry of the DNA. (c–e) Results show moving average values, averaged over 0.5 ns, from three explicit-solvent (TIP3P) PME and GB simulations. Connecting lines are shown to guide the eye. Images were rendered using VMD (100). To see this figure in color, go online.

The collapsed state was defined as tail extension of $<60 \text{ \AA}$, representing the value below which all tails have clearly collapsed, both in explicit and implicit solvents (Fig. 2).

On average, the H3' tail collapse is ~ 100 times faster in the implicit-solvent GB simulation ($0.10 \pm 0.05 \text{ ns}$) than in the explicit-solvent (TIP3P) PME simulation ($10 \pm 4 \text{ ns}$); the H2B' tail collapse is ~ 7 times faster ($0.05 \pm 0.02 \text{ ns}$ for GB compared to $0.34 \pm 0.05 \text{ ns}$ for PME), and the H3 tail collapse is almost no faster ($0.71 \pm 0.23 \text{ ns}$ for GB compared to $0.82 \pm 0.21 \text{ ns}$ for PME). Clearly, there can be large variations from simulation to simulation.

Notice that in Fig. 2 *a*, the tails of H2B' and H3 are not extended as far from the DNA as is the tail of H3'. Fig. 2 also shows that the H3' N-terminus starts off $\sim 10 \text{ \AA}$ farther from the DNA than the N-termini for H3 and H2B'. This suggests that the H3 and H2B' tails are already partially collapsed at the beginning of the simulation, which may explain why the speedup for the H3' tail collapse is much larger than for the H2B' and H3 tails.

Nucleosome DNA unwrapping

Accessibility to the DNA is critical for gene transcription, and therefore, the unwrapping of nucleosome DNA from the histone core has been extensively studied both experimentally (86) and computationally (87). Experiments have previously shown that at high pH, there is a loosening of the nucleosome structure (88). The implicit- and explicit-solvent TIP3P simulations show a similar effect, with the ends of the DNA unwrapping from the nucleosome core (Fig. 3). The ends of the DNA in an unwrapped state are extremely flexible, with fluctuations of $>10 \text{ \AA}$ in end-to-end DNA distance after unwrapping. However, the DNA unwraps on average ~ 30 times faster in the GB simulation ($0.04 \pm 0.02 \text{ ns}$) than in the explicit-solvent (TIP3P) PME simulation ($1.3 \pm 0.4 \text{ ns}$), averaged over three MD runs of equal length. Clearly, there can be large variations from simulation to simulation.

Note that even after 30 ns of explicit-solvent (TIP3P) PME simulation, the unwrapped DNA ends fail to fully extend out from the histone core, i.e., the end-to-end DNA distance in the explicit-solvent (TIP3P) PME simulation fails to exceed 150 \AA . This is in contrast to the GB simulation, where the end-to-end DNA distance exceeds 200 \AA after 2 ns of simulation. Most likely, the DNA in the explicit-solvent (TIP3P) PME simulation is unable to extend farther due to electrostatic repulsion between the ends of the DNA in the central cell and the DNA in neighboring cells, under the periodic boundary conditions. Therefore, to compare the GB and explicit-solvent (TIP3P) PME conformational sampling and effective speeds on an equal footing, we defined the unwrapped state as end-to-end DNA distance of 125 \AA , a state that is sampled by both simulations (Fig. 3, *a* and *b*). Fig. 3, *a* and *b*, shows that the unwrapping process proceeds along

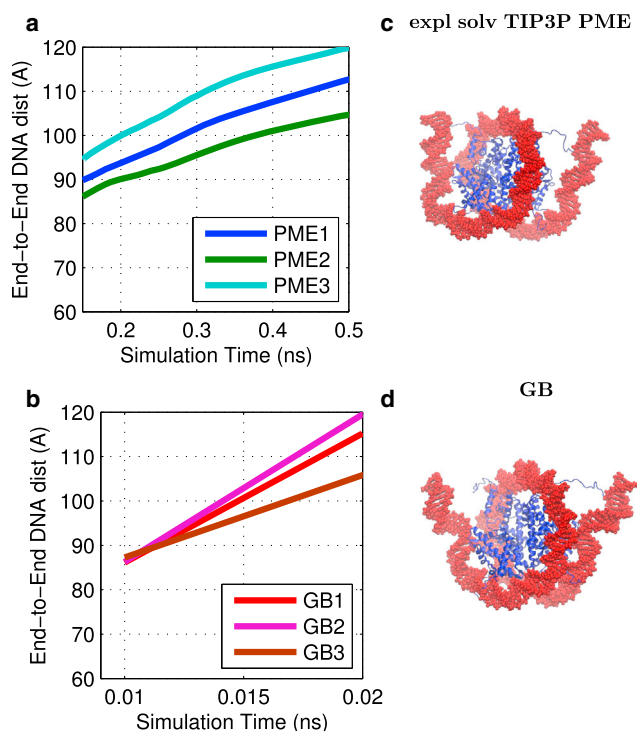


FIGURE 3 Nucleosome DNA unwrapping at high pH. (*a* and *b*) DNA end-to-end distance for three explicit-solvent (TIP3P) PME (*a*) and GB (*b*) simulations. Moving average values, averaged over 0.5 ns, are shown, with connecting lines to guide the eye. (*c*) Ends of the DNA unwrapped from the histone core after 0.8 ns of an explicit-solvent (TIP3P) PME simulation. (*d*) Ends of the DNA unwrapped from the histone core after 0.03 ns of a GB simulation. DNA surface is shown in red and histone backbone structure in blue. Images were rendered using VMD (100). To see this figure in color, go online.

similar reaction coordinates for both the PME and GB simulations, but over different timescales.

Methodological details

To study DNA unwrapping, we used the 1KX5 nucleosome with the histone tails removed (85), because the tails may hinder the unwrapping process. The condition of high pH was mimicked by setting all of the titratable sites to their deprotonated states. A larger box extending 36 \AA beyond the solvent was used for the explicit-solvent (TIP3P) PME simulation to allow room for the DNA to unwrap.

Mixed case

Folding of a miniprotein

Generally, the protein folding problem can be broken down (89) into three subproblems: 1) predicting the folded structure; 2) identifying the folding pathway(s); and 3) quantifying the forces that determine the folded state. We focus here specifically on subproblem 1, predicting the folded structure, where success can be clearly quantified. For the

exercise, we have chosen CLN025, which is arguably the smallest (only 10 residues) fast-folding miniprotein, previously used in atomistic studies of protein folding (70). The most straightforward approach would have been to compare the times it took for the protein to regain its known native conformation starting from a completely unfolded conformation at 300 K in explicit-solvent (TIP3P PME) versus implicit-solvent (GB) simulations. However, in contrast to implicit solvent, in which CLN025 readily folds on the timescale of several hundred nanoseconds, in explicit solvent at 300 K, not a single complete folding event occurred over 3 μ s of the simulation time—a realistic maximum in this work. Fortunately, the problem of kinetic traps at 300 K can be circumvented by a folding protocol (90–92) commonly used in studies of protein folding in both explicit and implicit solvent. In this protocol, the protein is simulated at its melting temperature, at which the native and unfolded conformations are supposed to be equally represented, and the issue of kinetic traps is mitigated.

The problem of identifying the correct native state from the MD trajectory is solved most easily for the GB simulation (Fig. 4 *d*) by choosing the snapshot corresponding to the absolute minimum potential energy (molecular mechanics energy plus free energy of solvation), which yields a structure that deviates from the correct native one by 1.4 Å. In explicit solvent, the total potential energy is dominated by the noise from the solvent box, and so the entire folding landscape is needed (Fig. 4 *c*) to confirm that the free-energy minimum is indeed close (1.5 Å RMSD) to the known native state. To proceed, we define folding time as the transition time from an unfolded structure to the next folded structure, following the approach in Lindorff-Larsen et al. (93). Based on the work of those authors, we defined the folded and unfolded states as structures with backbone RMSD values of <1.5 Å and >4.5 Å, respectively, consistent with the location of the folded and unfolded basins of the miniprotein (Fig. 4 *c*). At the experimental melting temperature of ~340 K, the GB simulation (Fig. 4 *b*) yields an average folding time of 10 ± 1 ns, compared to 70 ± 30 ns for the explicit-solvent (TIP3P) PME simulation (Fig. 4 *a*). The numbers of independent folding events are statistically significant in the two cases, 151 and 14, respectively.

Let us now touch upon the more complex folding subproblems, 2 and 3. At the experimental melting temperature, the folded and unfolded states should be equally occupied. However, it is obvious from Fig. 4 that, qualitatively, neither method gives the 1:1 ratio of the native folded to the completely unfolded states. In fact, at 340 K in the GB solvent model, the native folded state is occupied only ~1% of the time and the unfolded state ~40% of the time. For the explicit-solvent TIP3P PME simulation, the qualitative trend is the opposite: at the melting temperature (340 K), the unfolded state is undersampled (~3% of the time), whereas

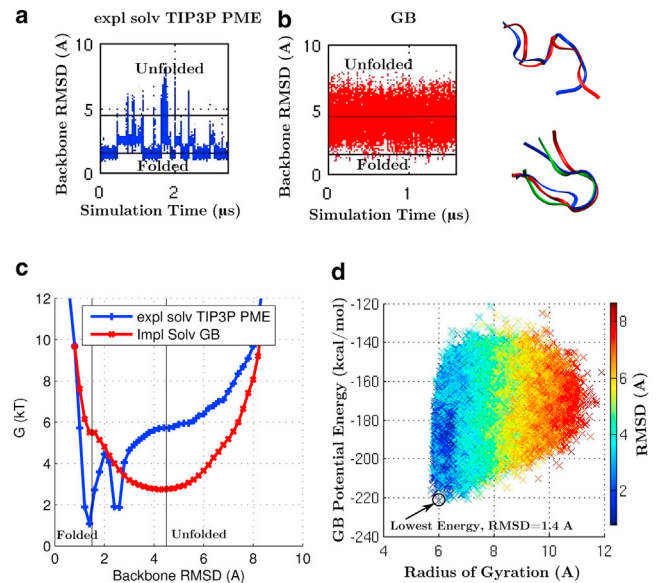


FIGURE 4 CLN025 miniprotein folding at its experimental melting temperature of 340 K. (*a* and *b*) RMSD of backbone heavy atoms relative to the starting structure for the explicit-solvent (TIP3P) PME simulation (*a*) and the GB simulation (*b*). The horizontal lines represent RMSD = 1.5 and 4.5 Å. Folded states are states with RMSD < 1.5 Å and unfolded states are states with RMSD > 4.5 Å. The trajectory is sampled every 100 ps for calculation of the RMSD values shown here. (*c*) Free-energy landscape for the explicit-solvent TIP3P PME and the GB simulations. (*d*) Potential energy, including solvation free energy, from the GB simulation, as a function of the distance (RMSD) from the experimental native structure. The lowest-energy structure approximates the correct folded state, as indicated by the low RMSD values. (*Inset*) Images of protein backbone conformations from representative snapshots for folded and unfolded states from the explicit-solvent (TIP3P) PME (*blue*) and GB (*red*) simulations. The starting structure (*green*) is shown for comparison. Images rendered using VMD (100). To see this figure in color, go online.

the native folded state is occupied ~50% of the time. In both cases, a large percentage of the states (~50% for both GB and PME) represent an intermediate state that is neither native folded nor completely unfolded. The Lindorff-Larsen et al. study (93) estimates that an explicit-solvent PME simulation at 370 K would be more representative of the melting temperature. The above suggests that the folded structure for CLN025 is under-stabilized in the GB model and over-stabilized in the explicit-solvent PME model, at least for the specific gas-phase force-field used (AMBER ff10). Another problem with the TIP3P PME simulation is that it yields a stable misfolded state at ~2.5 Å RMSD (Fig. 4 *c*), in addition to the correct native state at ~1.5 Å RMSD. The appearance of a nonnative compact intermediate is in contradiction to experiment, where CLN025 has been characterized as a two-state fast-folding protein with a unique native conformation (70). Given the obvious discrepancies between simulation and experiment, it is unclear how accurate the full free-energy profile of CLN025 produced by either of these simulations is, relative to experiment. Therefore, neither of the solvent models (with the underlying gas-phase force

field used) may be suitable to address the more difficult and detailed subproblem 2, identifying the correct folding pathway. We can still ask, however, what would it take for the GB model to reproduce the TIP3P PME folding landscape as faithfully as possible? Obviously, the landscapes at 340 K (Fig. 4 c) are very different. We found, however, that by lowering the GB simulation temperature, one can improve the agreement between the GB and TIP3P landscapes (see the Supporting Material). Specifically, at 260 K, the GB landscape becomes a fair approximation to the explicit-solvent one at 340 K (Fig. 4 c), especially near its native basin. Given the similarity of the landscapes, it makes sense to estimate how much faster is the GB simulation, as compared to the explicit solvent, in its ability to generate the landscape (i.e., the potential of mean force). Assuming that two similar-shaped landscapes of equivalent resolution require the same number of folding events for their construction, we find that the GB model at 260 K is around three times faster per nanosecond of simulation than the TIP3P PME model at 340 K at generating the free-energy landscape of CLN025.

Subproblem 3, quantifying the forces that determine the folded state, is an interesting case. Energy landscapes are easier to obtain and are more intuitive in the implicit-solvent model, in which the solvent degrees of freedom are averaged out; each snapshot has a definite energy, including the free energy of solvation. In that sense, the model can help with physical reasoning, as long as one is aware of its limitations. For example, investigating folding forces near the energy minimum representing the correctly predicted native state of CLN025 is an appropriate task for the implicit-solvent model.

Effect of Langevin dynamics collision frequency, γ , on the speed of conformational sampling

Langevin dynamics was used for temperature regulation in our simulations, where the equation of motion for a particle i is given by (81)

$$m_i \frac{\partial^2 x_i(t)}{\partial t^2} = F_i(t) - \gamma \frac{\partial x_i(t)}{\partial t} m_i + R_i(t), \quad (5)$$

where the particle has mass m_i , position x_i , and a force F_i acting on it at time t . A random force R_i represents the influence of the solvent or heat bath. R_i is assumed to be uncorrelated with the positions, velocities, and forces of the particle acting on it. In the regime $\gamma \delta t \ll 1$, typically used in biomolecular simulations, where δt is the simulation time step, R_i is taken from a Gaussian distribution with mean zero and variance,

$$\langle R_i^2(t) \rangle = 2m_i \gamma k_B T / \delta t, \quad (6)$$

where k_B is the Boltzmann constant and T is the absolute temperature. The preceding simulations used a collision

frequency of $\gamma = 0.01 \text{ ps}^{-1}$. A value of $\gamma \sim 50 \text{ ps}^{-1}$ would mimic a typical aqueous environment (94). However, lower values of γ are commonly used to increase the speed of conformational change in practical MD simulations. Fig. 5 shows the effect of different values of γ on the average speed of conformational sampling for GB simulations, for the conformational changes considered here: in general, conformational sampling speed decreases with increasing γ .

The collision frequency, γ , in the Langevin equation simultaneously controls two forces (62) expected to have opposing effects on conformational sampling speed: friction force (viscosity) and the random collision force—thermal kicks—due to Brownian motion (47,62,94). The latter may enhance conformational sampling by providing additional energy for crossing energy barriers (62). Thus, lowering γ reduces the beneficial effects of the thermal kicks on the sampling speed, but at the same time it helps speed up the sampling by increasing the diffusion rate. Which one of the two effects wins may depend on the type of the system and conformational change, as well as the value of γ . We find that for all the conformational changes considered here, lower values of γ always result in faster conformational sampling on average. That the speedups generally increase with decreasing collision frequency suggests that the speedup due to reduction in effective viscosity dominates any possible slowdown due to the decreased intensity of thermal kicks that help overcome small barriers in the energy landscape. We conjecture that the slowdown is relatively minor, because the free-energy landscapes in the GB model used here are relatively smooth,

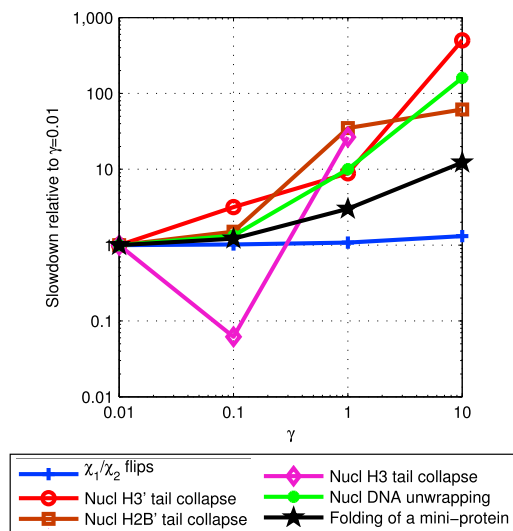


FIGURE 5 Effect of Langevin dynamics collision frequency, γ , on conformational sampling speed of GB simulations. Slowdown is computed relative to the average speed of conformational change at the baseline value of $\gamma = 0.01 \text{ ps}^{-1}$. The H3 histone tail did not collapse for $\gamma = 10 \text{ ps}^{-1}$, even after 22 ns of simulation (over 20 days of wall clock time), and therefore, this data point is not shown. Connecting lines shown to guide the eye. To see this figure in color, go online.

perhaps even artificially so compared to explicit solvent (53) and, possibly, reality (95).

The speedup in conformational sampling shown in Fig. 6 for GB with collision frequency $\gamma = 0.01 \text{ ps}^{-1}$ relative to TIP3P PME itself has two potential contributing factors, the associated reduction in solvent viscosity relative to the explicit solvent, and possible differences in the free-energy landscapes produced by the two solvent models (see, e.g., Fig. 4). To estimate the contribution to the entire conformational sampling speedup due to the decrease of the effective solvent viscosity alone, we note that alterations of the free-energy barriers, if they occur, are expected to be independent of the value of γ . With this in mind, the average viscosity-reduction speedup can be approximated from Fig. 5 as the increase of the sampling speed of $\gamma = 0.01 \text{ ps}^{-1}$ simulations relative to the $\gamma = 10.0 \text{ ps}^{-1}$ case, the latter being more representative of the collision frequency for liquid water ($\sim 50 \text{ ps}^{-1}$). The speedups are ~ 1 -fold, ~ 60 - to 500 -fold, and ~ 10 -fold for small (χ_1 and χ_2 dihedral angle flips), large (nucleosome histone tail collapse and nucleosome DNA unwrapping), and mixed (folding of a miniprotein) conformational changes, respectively. Comparing these viscosity-reduction speedups to the total speedups for the GB relative to explicit-solvent PME simulations, ~ 1 -fold, ~ 1 - to 100 -fold, and ~ 7 -fold for small, large, and mixed conformational changes, we see that the speedup due to lower collision frequency is, on average, not smaller than the entire conformational sampling speedup of GB compared to the explicit-solvent PME. Thus, relative to possible speedups due to alterations of the energy landscapes, for the systems studied here, the speedup due to the reduction in effective viscosity is dominant. The above analysis is necessarily very qualitative. For example,

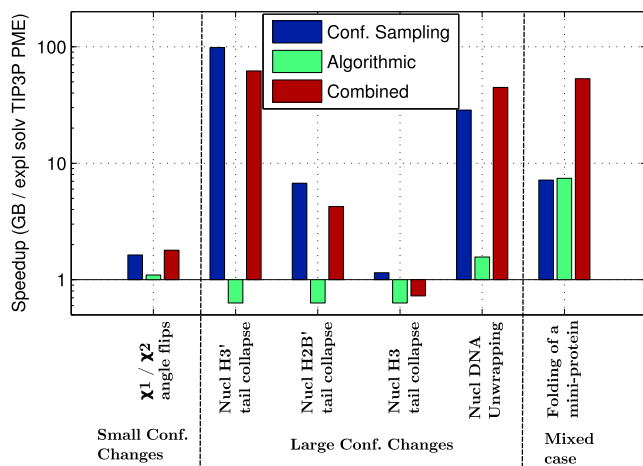


FIGURE 6 Average speedup for GB relative to explicit-solvent (TIP3P) PME simulations. The combined speedup includes the average speedup due to differences in the rate of conformational sampling, as well as due to algorithmic differences between the two methods. The Langevin dynamics collision frequency of $\gamma = 0.01 \text{ ps}^{-1}$ was used for these simulations. To see this figure in color, go online.

the collision frequency of 10.0 ps^{-1} , or even 50.0 ps^{-1} , only crudely approximates that of water. And TIP3P itself is ~ 2.4 times faster than real water as measured by its self-diffusion rate (96).

The combined speedup

Fig. 6 shows the combined speedup resulting from the use of implicit solvation (GB) relative to explicit treatment of the solvent (TIP3P + PME). The combined speedup comes from two distinct sources, 1) the increased rate of conformational sampling, and 2) algorithmic speedup.

The rate of conformational sampling is measured as the average simulation time required for a given conformational change. The effect of algorithmic speedup is measured by simulation time per unit of wall clock time (ns/day). The average combined speedup ranges from ~ 0.7 -fold to ~ 60 -fold for the structures and conformational changes studied here. Alternatively, one can think of this speedup in terms of an effective simulation time for GB MD. This effective time can be defined as the equivalent simulation time in explicit-solvent TIP3P PME MD (on the same resource) required to achieve comparable sampling of conformational space. For example, combining the $\sim 574 \text{ ns/day}$ nominal computation speed for GB with the approximately sevenfold average speedup of conformational sampling, results in $\sim 4 \mu\text{s/day}$ effective speed for GB (see Table 1). The corresponding production simulation run of one month could probe effective timescales of $> 100 \mu\text{s}$.

Without further approximations (97,98) aimed at improving the scaling (efficiency) of the model, the algorithmic complexity of GB scales as n^2 , whereas the scaling is $M \log N$ for explicit-solvent (TIP3P) PME, where n is the number of solute atoms and N is the total number of atoms in the system, including the solvent. Therefore, generally, algorithmic speedup decreases with structure size for GB relative to explicit-solvent (TIP3P) PME, since n^2 grows faster than $M \log N$, assuming that N itself does not grow faster than n^2 . This assumption is generally true for structures immersed in a solvent box of sufficient size to allow small conformational changes. However, in some cases, solvent boxes much larger than the solute are required to allow room for large conformational changes, such as for the unwrapping of DNA from the nucleosome core, or protein folding. In such cases, N becomes much larger than n , offsetting the PME advantage of scaling. Even for cases where explicit-solvent (TIP3P) PME is faster than GB (e.g., nucleosome tail collapse), when combined with the speedup due to conformational sampling, the overall speed of conformational sampling for GB can be faster.

Note that we did not use a cutoff for the calculation of long-range pairwise interactions for the GB simulations. Additional speedup of the GB simulation is possible using a cutoff. However, long-range electrostatic interactions decay slowly, and using a cutoff, which ignores such

TABLE 1 Speed of computation: explicit vs. implicit solvent

Simulation method	Structure and conformational change			
	Protein χ_1, χ_2 flips	Nucleosome tail collapse	Nucleosome DNA unwrap	Miniprotein folding
Explicit solvent TIP3P PME	20.55 \pm 0.02	1.79 \pm 0.02	0.92 \pm 0.02	74 \pm 1
GB (nominal speed)	21.86 \pm 0.26	1.13 \pm 0.00	1.39 \pm 0.006	574 \pm 2
GB (effective speed)	32.79 \pm 0.40	1.4 \pm 0.6/8 \pm 3/113 \pm 68	42 \pm 28	4018 \pm 1768

Nominal speed is the number of nanoseconds of simulation time per day of wall clock time (ns/day). Effective speed = (nominal speed) \times (conformational sampling speedup). Qualitatively, the effective speed is the estimated simulation time required by the explicit-solvent TIP3P PME simulation to sample conformational space comparable to the corresponding GB simulation run for 1 day on the same resource. The three time values for the nucleosome tail collapse represent the results for the collapse of the H3, H2B', and H3' tails, respectively. Langevin collision frequency is $\gamma = 0.01 \text{ ps}^{-1}$; AMBER-12 on a single GTX680 GPU card with the SPFP precision model.

interactions beyond the cutoff distance, can significantly reduce the accuracy of the GB calculation. The loss of accuracy can be particularly significant for large, highly charged structures, such as the nucleosome. Fortunately, GB-specific alternatives to cutoffs have been developed (97,98).

CONCLUSIONS

Atomistic molecular dynamics simulations are routinely used to study structure, function, and activity of biological molecules. However, the utility of such studies is often limited by inadequate sampling of conformational space. Implicit-solvent-based simulations can speed up the sampling of conformational space relative to explicit-solvent simulations, but the speedup comes at the cost of making additional approximations to reality. In particular, implicit-solvent simulations can sample conformational space faster than can explicit-solvent simulations, but they may also alter the free-energy landscapes. These issues are discussed on a case-by-case basis for the types of problems considered here. In practice, the choice of solvent model is dictated by the accuracy/speed trade-offs, which are system-dependent: knowing what to expect is critical to making the right choice.

In this study, we compared the speeds of small, large, and mixed conformational changes simulated in implicit solvent using the GB approximation to those simulated in explicit solvent using the PME model to approximate the long-range forces. The conformational changes studied here consist of dihedral angle flips, nucleosome histone tail collapse, DNA unwrapping from the nucleosome histone core, and mini-protein folding.

Our results show that speedup of conformational change in implicit solvent can vary considerably, depending on the details of the transition, and can range from no speedup at all to almost a 100-fold speedup. In general, the larger the conformational change, the higher the speedup one may expect from the use of the implicit rather than the explicit solvent, but this tendency is not universal or uniform. Within each type of transition, a range of speedups were observed: for example, side-chain dihedral angle flips in a 300-residue protein are, on average, only ~ 1.5 -fold faster in implicit solvent, but for some groups the ratio can be 10:1 or <1 .

We expect these speedup values to be specific to current AMBER GB flavors that offer a reasonable compromise between accuracy and speed. The use of GB models available in other packages may result in different speedups (47). At the same time, more general conclusions, discussed below, are expected to be less sensitive to the specifics of the GB model.

The free-energy landscapes for the explicit-solvent PME and implicit-solvent GB simulations at a given temperature can also be quite different, as seen in the case of mini-protein folding. Where the goal of the GB simulation is to reproduce a specific free-energy landscape in explicit solvent, e.g., by lowering the simulation temperature, the speedups noted above may be different. For example, the speedup in folding rate is considerably lower (approximately threefold) when comparing simulations with similar free-energy landscapes than when comparing simulations at the same temperature (speedup is approximately sevenfold) at 340K, the experimental melting temperature of the miniprotein examined here.

We also examined the effect of the Langevin dynamics collision frequency, γ , on the speed of conformational change. Decreasing the value of γ may have opposing effects on conformational sampling by simultaneously reducing the solvent viscosity and at the same time reducing random collision force, which can speed up sampling by providing additional energy for crossing small energy barriers. We find that for all systems and types of conformational changes considered here, increasing the collision frequency reduces the speed of conformational change. Thus, using low values of γ seems always of benefit to implicit-solvent simulations where enhanced sampling is desired, at least when the possibility of altered dynamics (99) or kinetics (51) is irrelevant or can be ignored. By analyzing how the speedup changes with increasing γ , from near zero to values representative of true viscosity of water, we have concluded that conformational sampling speedup is mainly due to reduction in solvent viscosity rather than to possible differences in free-energy landscapes between the implicit GB and explicit PME treatment of solvation.

The nominal computational speed of an implicit-solvent simulation, that is the number of nanoseconds per day of

simulation time, can be higher or lower than that of the corresponding explicit-solvent simulation due to differences in how the two approximations scale with system size. The most common explicit-solvent PME approximation for long-range forces scales as $\sim \log N$, whereas the standard (without further approximations) GB implicit computation scales as $\sim n^2$, where N is the total number of solvent and solute atoms combined and n is the number of solute atoms only. For example, a GB simulation of the 25,100-atom nucleosome is nominally 1.6 times slower than the explicit-solvent PME simulation with a small TIP3P solvent box extending 10 Å from the solute, whereas it is 1.6 times faster compared to a PME simulation in a larger 36 Å TIP3P solvent box. The choice of the box size is determined by the specifics of the problem, e.g., whether large conformational changes are expected. However, what matters in most cases is the effective speedup, which takes into account both computational speed and conformational sampling speed. For the types of conformational changes considered here, the effective speedup ranges from ~ 0.7 - to ~ 60 -fold for implicit-solvent GB simulations compared to explicit-solvent TIP3P PME simulations. For example, at 574 ns/day, the GB simulation for the 166-atom CLN025 miniprotein is approximately sevenfold faster than the explicit-solvent PME simulation (74 ns/day) on our computational platform. Combined with the additional approximately sevenfold conformational speedup, the 574 ns/day implicit-solvent MD simulation becomes equivalent to an explicit-solvent PME simulation at ~ 4 μ s/day for the CLN025 miniprotein folding simulation, making effective timescales of hundreds of microseconds accessible on conventional computing platforms. This example illustrates an important point: when it comes to sampling conformation space, simply considering computational speed, i.e., nanoseconds per day, may not be sufficient. Instead, one must also take into consideration the speed of conformational change when comparing implicit- and explicit-solvent methods. In this respect, the effective speedup, which takes into account both conformational search speed and computational speed, is more informative.

SUPPORTING MATERIAL

Supporting Materials and Methods, Supporting Results and Discussion, and seven figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(15\)00003-X](http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00003-X).

ACKNOWLEDGMENTS

The authors acknowledge use of the HokieSpeed supercomputer at Virginia Tech.

This work was supported by National Institutes of Health grant GM076121 to A.O., Burroughs-Wellcome fund grant 1011306 to R.A., National Science Foundation SI2-SSE grants NSF-1047875 and NSF-1148276 to R.C.W., a University of California UC Lab 09-LR-06-117793 grant to R.C.W., a fellowship from NVIDIA, and National Science Foundation grant CNS-0960081.

REFERENCES

1. Dodson, G. G., D. P. Lane, and C. S. Verma. 2008. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Rep.* 9:144–150.
2. Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA.* 102:6679–6685.
3. Dror, R. O., R. M. Dirks, ..., D. E. Shaw. 2012. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* 41:429–452.
4. Wang, W., O. Donini, ..., P. A. Kollman. 2001. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* 30:211–243.
5. Schlick, T., R. Collepardo-Guevara, ..., X. Xiao. 2011. Biomolecular modeling and simulation: a field coming of age. *Q. Rev. Biophys.* 44:191–228.
6. Shaw, D. E., M. M. Deneroff, ..., S. C. Wang. 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* 51:91–97.
7. Sagui, C., and T. A. Darden. 1999. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* 28:155–179.
8. Ratner, M. A. 2001. Biomolecular processes in the fast lane. *Proc. Natl. Acad. Sci. USA.* 98:387–389.
9. Grossfield, A., and D. M. Zuckerman. 2009. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 5:23–48.
10. Onufriev, A. 2010. Continuum electrostatics solvent modeling with the generalized Born model. In *Modeling Solvent Environments*. M. Feig, editor. Wiley, New York, pp. 127–165.
11. Schlick, T. 2002. *Molecular Modeling and Simulation. An Interdisciplinary Guide*. Springer-Verlag, New York.
12. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
13. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
14. Toukmaji, A. Y., and J. A. Board, Jr. 1996. Ewald summation techniques in perspective: a survey. *Comput. Phys. Commun.* 95:73–92.
15. York, D., and W. Yang. 1994. The fast Fourier Poisson method for calculating Ewald sums. *J. Chem. Phys.* 101:3298–3300.
16. Onufriev, A., D. Bashford, and D. Case. 2000. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B.* 104:3712–3720.
17. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins.* 55:383–394.
18. Mongan, J., C. Simmerling, ..., A. Onufriev. 2007. Generalized Born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* 3:156–169.
19. Still, W. C., A. Tempczyk, ..., T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
20. Bashford, D., and D. A. Case. 2000. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* 51:129–152.
21. Hawkins, G. D., C. J. Cramer, and D. G. Truhlar. 1995. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* 246:122–129.
22. Hawkins, G. D., C. J. Cramer, and D. G. Truhlar. 1996. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100:19824–19839.

23. Ghosh, A., C. S. Rapp, and R. A. Friesner. 1998. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B.* 102:10983–10990.
24. Lee, M. S., J. F. R. Salsbury, and C. L. Brooks, III. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116:10606–10614.
25. Tsui, V., and D. Case. 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122:2489–2498.
26. Cramer, C. J., and D. G. Truhlar. 1999. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* 99:2161–2200.
27. David, L., R. Luo, and M. K. Gilson. 2000. Comparison of generalized Born and Poisson models: energetics and dynamics of HIV protease. *J. Comput. Chem.* 21:295–309.
28. Im, W., M. S. Lee, and C. L. Brooks, 3rd. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–1702.
29. Schaefer, M., and M. Karplus. 1996. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* 100:1578–1599.
30. Calimet, N., M. Schaefer, and T. Simonson. 2001. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins.* 45:144–158.
31. Feig, M., W. Im, and C. L. Brooks, 3rd. 2004. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* 120:903–911.
32. Archontis, G., and T. Simonson. 2005. A residue-pairwise generalized born scheme suitable for protein design calculations. *J. Phys. Chem. B.* 109:22667–22673.
33. Feig, M., A. Onufriev, ..., C. L. Brooks, 3rd. 2004. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* 25:265–284.
34. Nymeyer, H., and A. E. García. 2003. Simulation of the folding equilibrium of α -helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA.* 100:13934–13939.
35. Scarsi, M., J. Apostolakis, and A. Caffisch. 1997. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A.* 101:8098–8106.
36. Dominy, B. N., and C. L. Brooks. 1999. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B.* 103:3765–3773.
37. Gallicchio, E., and R. M. Levy. 2004. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25:479–499.
38. Grant, J. A., B. T. Pickup, ..., A. Nicholls. 2007. The Gaussian generalized Born model: application to small molecules. *Phys. Chem. Chem. Phys.* 9:4913–4922.
39. Haberthür, U., and A. Caffisch. 2008. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* 29:701–715.
40. Spassov, V. Z., L. Yan, and S. Szalma. 2002. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B.* 106:8726–8738.
41. Ulmschneider, M. B., J. P. Ulmschneider, ..., A. Di Nola. 2007. A generalized Born implicit-membrane representation compared to experimental insertion free energies. *Biophys. J.* 92:2338–2349.
42. Tanizaki, S., and M. Feig. 2005. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* 122:124706.
43. Zhang, L. Y., E. Gallicchio, ..., R. M. Levy. 2001. Solvent models for protein-ligand binding: comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. *J. Comput. Chem.* 22:591–607.
44. Srinivasan, J., M. W. Trevathan, ..., D. A. Case. 1999. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* 101:426–434.
45. Zuckerman, D. M. 2011. Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.* 40:41–62.
46. Pierce, L. C., R. Salomon-Ferrer, ..., R. C. Walker. 2012. Routine access to millisecond timescale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* 8:2997–3002.
47. Feig, M. 2007. Kinetics from implicit solvent simulations of biomolecules as a function of viscosity. *J. Chem. Theory Comput.* 3:1734–1748.
48. Amaro, R. E., X. Cheng, ..., J. A. McCammon. 2009. Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized Born molecular dynamics and end-point free energy calculations. *J. Am. Chem. Soc.* 131:4702–4709.
49. Amaro, R. E., R. V. Swift, ..., R. M. Bush. 2011. Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.* 2:388.
50. Zagrovic, B., C. D. Snow, ..., V. S. Pande. 2002. Simulation of folding of a small α -helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* 323:927–937.
51. Zagrovic, B., and V. Pande. 2003. Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem.* 24:1432–1436.
52. Ruscio, J. Z., and A. Onufriev. 2006. A computational study of nucleosomal DNA flexibility. *Biophys. J.* 91:4121–4132.
53. Johnson, Q., U. Doshi, ..., D. Hamelberg. 2010. Water's contribution to the energetic roughness from peptide dynamics. *J. Chem. Theory Comput.* 6:2591–2597.
54. Hamelberg, D., T. Shen, and J. A. McCammon. 2006. Insight into the role of hydration on protein dynamics. *J. Chem. Phys.* 125:094905.
55. Tan, C., L. Yang, and R. Luo. 2006. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *J. Phys. Chem. B.* 110:18680–18687.
56. Baker, N. A. 2005. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.* 15:137–143.
57. Zhou, R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins.* 53:148–161.
58. Beauchamp, K. A., Y.-S. Lin, ..., V. S. Pande. 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* 8:1409–1414.
59. Hu, Z., and J. Jiang. 2010. Assessment of biomolecular force fields for molecular dynamics simulations in a protein crystal. *J. Comput. Chem.* 31:371–380.
60. Shirts, M. R., and V. S. Pande. 2005. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* 122:134508.
61. Hagen, S. J., L. Qiu, and S. A. Pabit. 2005. Diffusional limits to the speed of protein folding: fact or friction? *J. Phys. Condens. Matter.* 17:S1503–S1514.
62. Jagielska, A., and H. A. Scheraga. 2007. Influence of temperature, friction, and random forces on folding of the B-domain of staphylococcal protein A: all-atom molecular dynamics in implicit solvent. *J. Comput. Chem.* 28:1068–1082.
63. Gee, P. J., and W. F. van Gunsteren. 2005. Numerical simulation of the effect of solvent viscosity on the motions of a β -peptide heptamer. *Chemistry.* 12:72–75.
64. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
65. Okur, A., D. R. Roe, ..., C. Simmerling. 2007. Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J. Chem. Theory Comput.* 3:557–568.
66. Roitberg, A. E., A. Okur, and C. Simmerling. 2007. Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. *J. Phys. Chem. B.* 111:2415–2418.
67. Grubmüller, H., H. Heller, ..., K. Schulten. 1991. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simul.* 6:121–142.

68. Watanabe, M., and M. Karplus. 1993. Dynamics of molecules with internal degrees of freedom by multiple timestep methods. *J. Chem. Phys.* 99:8063–8074.
69. Tuckerman, M. E., G. J. Martyna, and B. J. Berne. 1990. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.* 93:1287–1291.
70. Honda, S., T. Akiba, ..., K. Harata. 2008. Crystal structure of a ten-amino acid protein. *J. Am. Chem. Soc.* 130:15327–15331.
71. Honda, S., K. Yamasaki, ..., H. Morii. 2004. 10 residue folded peptide designed by segment statistics. *Structure.* 12:1507–1518.
72. Berman, H. M., J. Westbrook, ..., P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
73. Heinz, D. W., M. Ryan, ..., O. H. Griffith. 1996. Crystal structure of phosphatidylinositol-specific phospholipase C from *Bacillus cereus* in complex with glucosaminyl $_{\alpha 1 \rightarrow 6}$ -D-myo-inositol, an essential fragment of GPI anchors. *Biochemistry.* 35:9496–9504.
74. Davey, C. A., D. F. Sargent, ..., T. J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319:1097–1113.
75. Anandakrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H⁺⁺ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40:W537–W541.
76. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
77. Wickstrom, L., A. Okur, and C. Simmerling. 2009. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys. J.* 97:853–856.
78. Götz, A. W., M. J. Williamson, ..., R. C. Walker. 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8:1542–1555.
79. Salomon-Ferrer, R., A. W. Götz, ..., R. C. Walker. 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9:3878–3888.
80. Grand, S., A. Goetz, and R. Walker. 2013. SPFP: Speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* 184:374–380.
81. Pastor, R. W., B. R. Brooks, and A. Szabo. 1988. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65:1409–1419.
82. Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23:327–341.
83. Bertin, A., A. Leforestier, ..., F. Livolant. 2004. Role of histone tails in the conformation and interactions of nucleosome core particles. *Biochemistry.* 43:4773–4780.
84. Ferreira, H., J. Somers, ..., T. Owen-Hughes. 2007. Histone tails and the H3 α N helix regulate nucleosome mobility and stability. *Mol. Cell. Biol.* 27:4037–4048.
85. Biswas, M., K. Voltz, ..., J. Langowski. 2011. Role of histone tails in structural stability of the nucleosome. *PLOS Comput. Biol.* 7:e1002279.
86. Thåström, A., P. T. Lowary, and J. Widom. 2004. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods.* 33:33–44.
87. Voltz, K., J. Trylska, ..., J. Langowski. 2012. Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study. *Biophys. J.* 102:849–858.
88. Libertini, L. J., and E. W. Small. 1984. Effects of pH on the stability of chromatin core particles. *Nucleic Acids Res.* 12:4351–4359.
89. Dill, K. A., S. B. Ozkan, ..., T. R. Weikl. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:289–316.
90. Simmerling, C., B. Strockbine, and A. E. Roitberg. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124:11258–11259.
91. Kannan, S., and M. Zacharias. 2009. Simulated annealing coupled replica exchange molecular dynamics—an efficient conformational sampling method. *J. Struct. Biol.* 166:288–294.
92. Snow, C. D., E. J. Sorin, ..., V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69.
93. Lindorff-Larsen, K., S. Piana, ..., D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.
94. Loncharich, R. J., B. R. Brooks, and R. W. Pastor. 1992. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers.* 32:523–535.
95. Milanesi, L., J. P. Waltho, ..., M. Volk. 2012. Measurement of energy landscape roughness of folded and unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 109:19563–19568.
96. Izadi, S., R. Anandakrishnan, and A. V. Onufriev. 2014. Building water models: a different approach. *J Phys Chem Lett.* 5:3863–3871.
97. Anandakrishnan, R., M. Daga, and A. V. Onufriev. 2011. An *n* log *n* Generalized Born approximation. *J. Chem. Theory Comput.* 7:544–559.
98. Anandakrishnan, R., and A. V. Onufriev. 2010. An *N* log *N* approximation based on the natural organization of biomolecules for speeding up the computation of long range interactions. *J. Comput. Chem.* 31:691–706.
99. Rhee, Y. M., and V. S. Pande. 2008. Solvent viscosity dependence of the protein folding dynamics. *J. Phys. Chem. B.* 112:6221–6227.
100. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations

Supporting Material

Ramu Anandakrishnan,[†] Aleksander Drozdetski,[‡], Ross Walker,[§] and Alexey V. Onufriev^{† ‡}

[†] Department of Computer Science, Virginia Tech, Blacksburg, Virginia; [‡] Department of Physics, Virginia Tech, Blacksburg, Virginia; [§] San Diego Supercomputer Center, University of California, San Diego, California;

Methods

The use of effective Born radius as a metric of atom burial.

By definition, the effective Born radius B (see Eq. (4)) of a charge (atom) is inversely proportional to the desolvation penalty of taking it from water into the low dielectric interior of the protein, in the absence of all other charges. Thus, in general, one expects the degree of atom's burial δ beneath the molecular surface to correlate with its effective Born radius. In fact (1), for a perfect sphere of radius A inside a high dielectric solvent, the effective radius $B = A(1 - (\delta - A)^2/A^2)$, which increases monotonically with δ . Near the surface, R is directly proportional to the degree of burial. For non-spherical shapes and deeply buried groups the relationship is less linear and more of an approximation.

Simulation protocol

The following protocol was used for MD simulations, unless otherwise stated in the text. The simulations consisted of five stages – minimization, heating, two equilibration stages and the production stage. The Amber12 script for each of these stages is shown below. “xxx” in the following refers to the number of solute groups.

Minimization - PME

```
&cntrl
  imin=1,           ! minimize
  maxcyc=2000,      ! number of minimization steps
  ntb=1,           ! with periodic box
  igb=0,           ! pme
  cut=8,           ! cutoff for electrostatics
  ntr=1,           ! use restraints
  ntx=1,           ! input: formatted coord only
  irest=0,         ! input: not a restart file
  ntrpr=10,        ! output: print every 10 steps
  ntwx=0,          ! output: no trajectory
  ntwr=500000,     ! output: restart file written at the end
  iwrap=0,         ! output: do not wrap coord in restart/traj (for pme)
/
```

RESTRAIN ALL

5.0

RES 1 xxx

END

Minimization - GB

```

&cntrl
  imin=1,          ! minimize
  maxcyc=2000,    ! run for n steps of minimization
  ntb=0,          ! no box
  igb=5,          ! OBC gb
  cut=9999.0,     ! no cutoff for electrostatics
  rgbmax=15.0,    ! gb: cutoff for Born radii
  saltcon=0.145, ! gb: salt concentration
  gbsa=0,        ! gb: do not include surface area term
  ntr=1,         ! use restraints
  ntx=1,         ! input: formatted coord only
  irect=0,       ! input: not a restart
  ntp=10,        ! output: print every 10 steps
  ntwe=0,        ! output: do not write energy and temperature file
  ntwx=0,        ! output: no trajectory
  ntwprt=0,      ! output: all atoms written to traj
  ntwr=500000,  ! output: restart file written at the end
/
RESTRAIN ALL
5.0
RES 1 9999
END

```

Heating - PME

```

&cntrl
  imin=0,          ! MD, not minimization
  ig=-1,          ! random seed based on date/time
  nstlim=300000,  ! number of steps
  dt=0.002,       ! time step (ps)
  ntc=2,          ! shake: constrain H bonds
  ntf=2,          ! shake: ignore H bond interactions
  ntt=3,          ! 3 = langevin dynamics
  gamma_ln=0.01, ! collision frequency for langevin dynamics
  tempi=0.0,       ! initial temperature
  temp0=300.0,    ! reference temperature
  ntp=0,          ! no pressure scaling (const volume)
  ntb=1,          ! periodic boundary for const volume md
  igb=0,          ! pme
  cut=8.0,        ! cutoff for electrostatics
  ntr=1,         ! use restraints
  ntx=1,         ! input: formatted coord only
  irect=0,       ! input: not a restart
  ntp=500,        ! output: print every 500 steps
  ntwx=0,        ! output: no trajectory
  ntwprt=0,      ! output: all atoms written to traj
  ntwr=500000,  ! output: restart file written at the end
  iwrap=0,       ! output: do not wrap coord in restart/traj (for pme)
/
RESTRAIN ALL
1.0
RES 1 xxx
END

```

Heating - GB

```

&cntrl

```

```

imin=0,          ! MD, not minimization
ig=-1,          ! random seed based on date/time
nstlim=300000,  ! number of steps
dt=0.002,       ! time step (ps)
ntc=2,          ! shake: constrain H bonds
ntf=2,          ! shake: ignore H bond interactions
ntt=3,          ! 3 = langevin dynamics
gamma_ln=0.01, ! collision frequency for langevin dynamics
tempi=0.0,      ! initial temperature
temp0=300.0,    ! reference temperature
ntp=0,          ! no pressure scaling (const volume)
ntb=0,          ! no box
igb=5,          ! OBC gb
cut=9999.0,     ! no cutoff for electrostatics
rgbmax=15.0,    ! gb: cutoff for Born radii
saltcon=0.145, ! gb: salt concentration
gbsa=0,         ! gb: include surface area term
ntr=1,          ! use restraints
ntx=1,          ! input: formatted coord only
irest=0,        ! input: not a restart
ntpr=50,        ! output: print every n steps
ntwe=0,         ! output: do not write energy and temperature file
ntwx=0,         ! output: no trajectory
ntwprt=0,       ! output: n atoms written to traj (0=all atoms)
ntwr=500000,   ! output: restart file written at the end
/
RESTRAIN ALL
1.0
RES 1 9999
END

Equilibration 1 PME - Only differences from Heating are shown here
  nstlim=1000000, ! number of steps
  tempi=300.0,     ! initial temperature
  ntp=1,          ! isotropic pressure scaling (const pressure)
  ntb=2,          ! periodic boundary condition for const pressure md
/
RESTRAIN ALL
0.1

Equilibration 1 GB - Only differences from Heating are shown here
  nstlim=1000000, ! number of steps
  tempi=300.0,     ! initial temperature
/
RESTRAIN ALL
0.1

Equilibration 2 - Only differences from Equilibration 1 are shown here
RESTRAIN ALL
0.01

Production - Only differences from Equilibration 2 are shown here
  nstlim=100000000, ! number of steps
  ntr=0,            ! no restraints

```

Calculation of standard error

Standard error se of the mean is calculated as $se = \sigma/\sqrt{n}$, where σ is the standard deviation and n is the number of samples. For the product or ratio of two means, the standard error in the resulting product ($f = m_1 m_2$) or ratio ($f = m_1/m_2$) is calculated as:

$$se_f = f \sqrt{\left(\frac{se_1}{m_1}\right)^2 + \left(\frac{se_2}{m_2}\right)^2} \quad (1)$$

where se_f is the standard error of the resulting product or ratio of the mean values m_1 and m_2 with standard errors of se_1 and se_2 respectively.

Results and discussion

Analysis of dihedral angle flips

We analyzed the relationship between frequency of dihedral angle flips and (1) depth of burial within the protein as measured by effective Born radii, (2) extent of hydrogen bonding between side chains and water as measured by percent of samples with such hydrogen bonds, (3) difference in side-chain conformation as measured by residue RMSD, and (4) type of side-chain. However, the correlations in each of these cases were not statistically significant. Figure 1 shows that there is little correlation between residue burial depth and the difference between GB and PME χ_1/χ_2 dihedral angle flips – correlation coefficient < 0.5 . Figure 2 shows that there is little or no correlation between the difference in χ_1 and χ_2 dihedral flips and residue H-bonds

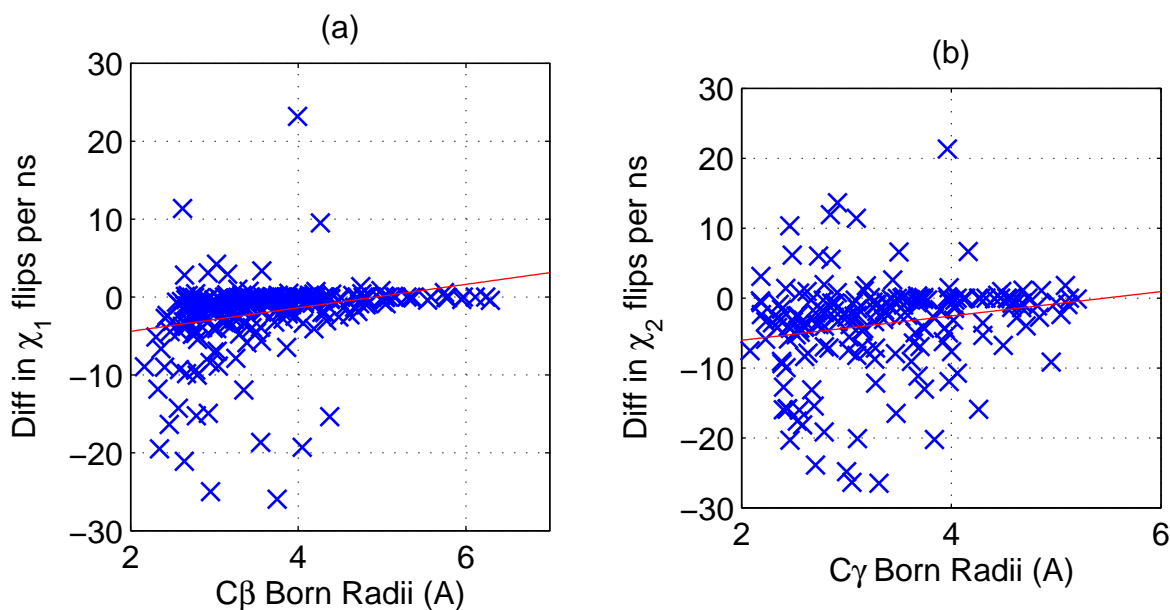


Figure 1: Difference in frequency of χ_1 and χ_2 dihedral angle flips as a function of residue burial depth. Depth of burial is measured as the Born radius of a residue's C_β atom for (a) χ_1 angles, and as the C_γ atom for (b) χ_2 angles.

(correlation coefficient < 0.1). Figure 3 shows that there is little correlation between side chain flexibility as measured by per residue RMS difference, and χ_1 or χ_2 dihedral angle flips (correlation coefficient < 0.5). Figure 4 shows that, excluding outliers, on average there is no significant difference in the frequency of χ_1/χ_2 dihedral angle flips.

For the 770 ns simulations considered here, the GB simulation explores different conformations than the explicit solvent (TIP3P) PME simulation. Figure 5 shows that the backbone RMSD relative to the starting structure is 1.6 Å for the explicit solvent (TIP3P) PME simulation compared to 4.5 Å for the GB simulation. Therefore to compare the χ_1 and χ_2 angles for the two simulations on an equal footing, we only considered groups where the distribution of the χ_1 and χ_2 angles were similar for the GB and explicit solvent (TIP3P) PME simulations, i.e. the frequency at which the dihedral angle ranges were sampled, differ by less than 10%. Figure 6 shows an example where the χ_1 and χ_2 angles sampled by the explicit solvent (TIP3P) PME and GB simulations are similar.

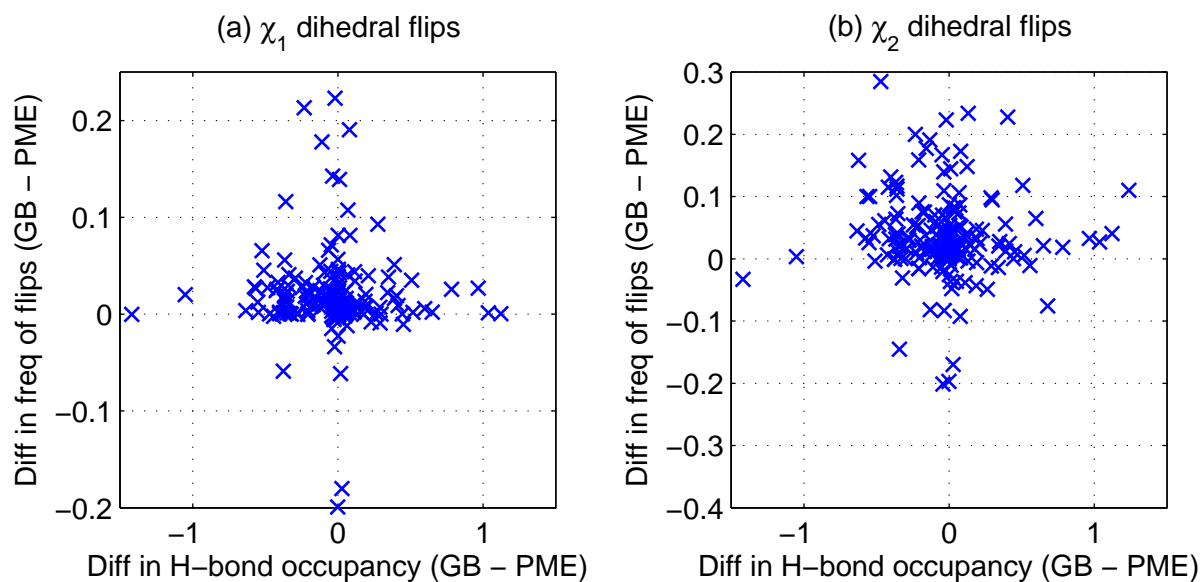


Figure 2: Difference (GB - PME) in χ_1 and χ_2 dihedral angle flips as a function of difference in H-bond occupancy.

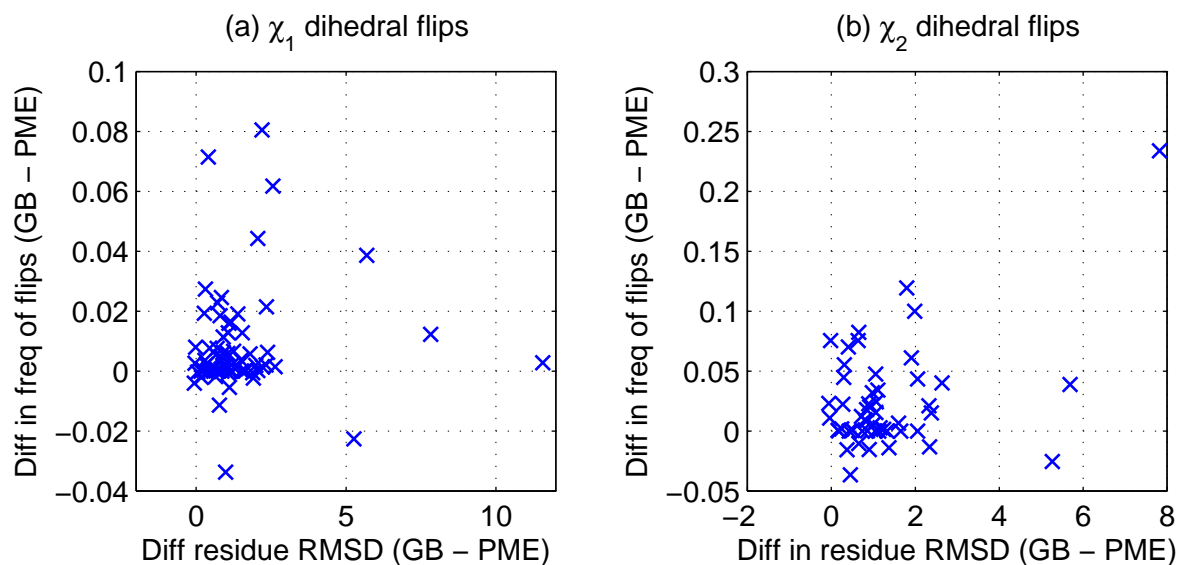


Figure 3: Difference in average frequency of χ_1 dihedral angle flips as a function of per residue RMS difference. RMS difference is measured relative to the backbone heavy atoms from the starting structure.

Free energy landscape temperature dependence

The CLN025 mini-protein folding simulations were run at the experimental melting temperature of 340 K. At this temperature the folded and unfolded states are expected to be sampled equally. However, due to limitations of the force fields and solvation models used, neither the explicit solvent TIP3P PME, nor the implicit solvent GB simulation equally sample the folded and unfolded states at the experimental melting temperature. For the GB model 260 K is more representative of the melting temperature, where the folded and unfolded states are sampled approximately equally (~ 11 and 15% respectively) (Fig. 7)

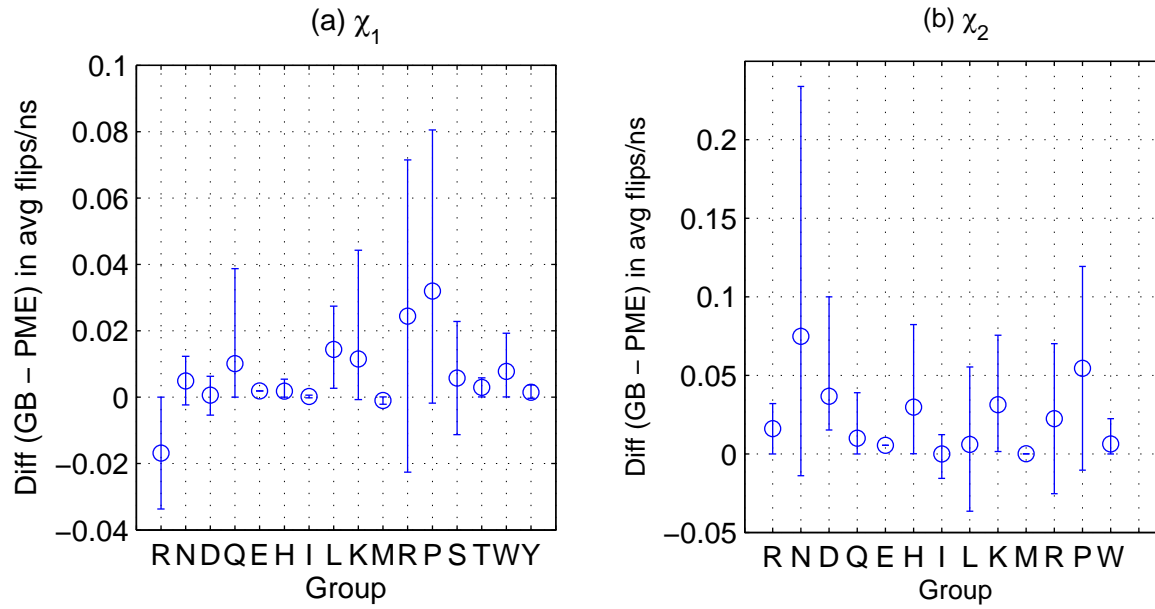


Figure 4: Difference (GB - PME) in average frequency of χ_1 and χ_2 dihedral angle flips as a function of residue type. Error bars indicate minimum and maximum values.

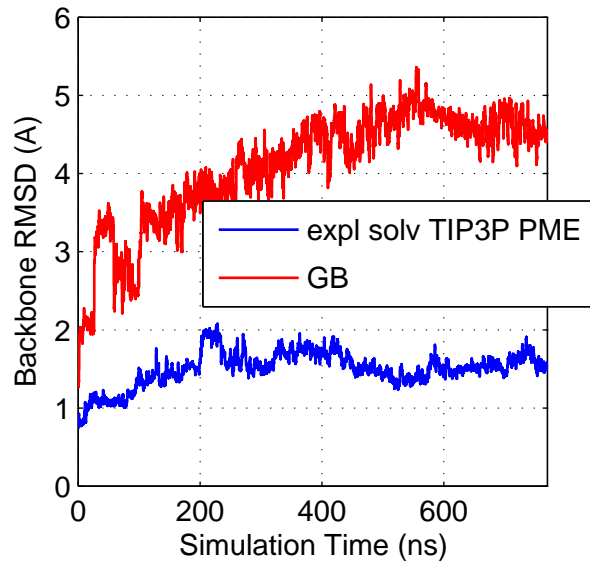


Figure 5: 770 ns simulations of 1GYM using the explicit solvent (TIP3P) PME and GB methods. Root mean square difference (RMSD) of backbone heavy atoms is relative to the starting structure. Figure shows moving average values averaged over 0.5 ns, with connecting lines shown to guide the eye.

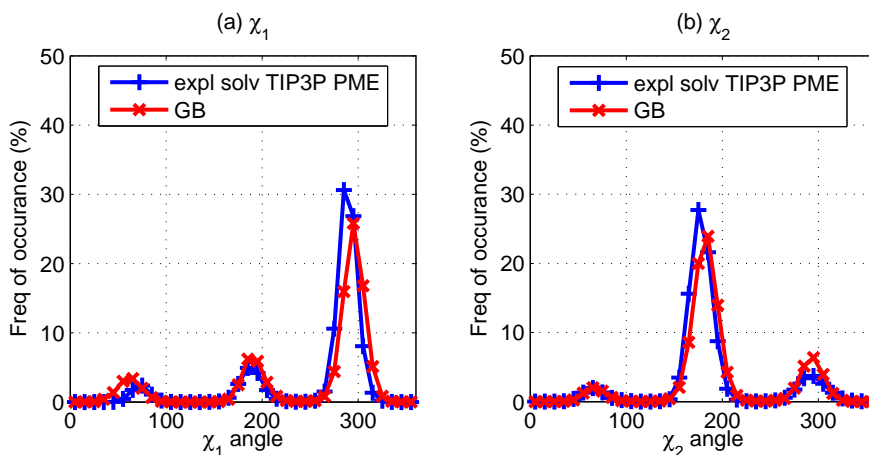


Figure 6: χ_1 and χ_2 angles sampled for residue GLN262 from the explicit solvent (TIP3P) PME and GB simulations of 1GYM. Both the explicit solvent (TIP3P) PME and GB simulations sample all three ranges of χ_1 and χ_2 angles, 0:120, 120:240, and 240:360. For calculating sampling frequency, the angles are grouped into 36 bins of 10 degrees each. Connecting lines are shown to guide the eye.

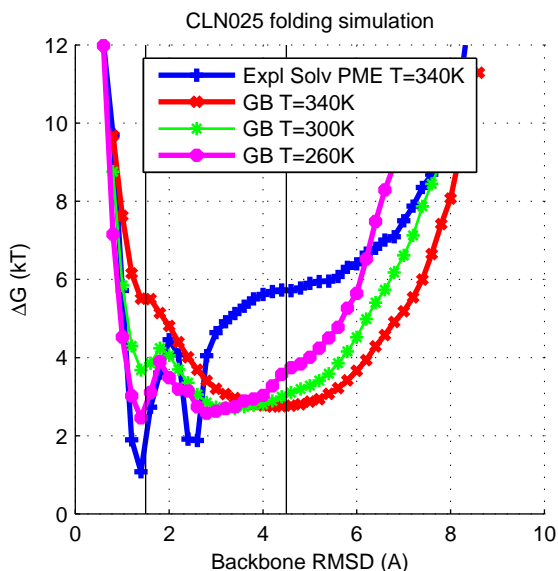


Figure 7: CLN025 mini-protein folding: Dependence of free energy landscape on temperature. GB simulations at T=340, 300 and 260K are shown with the explicit solvent TIP3P PME simulation at T=340K included for comparison. The horizontal lines represent RMSD = 1.5 and 4.5 Å. Folded states are states with RMSD < 1.5 Å and unfolded states are states with RMSD > 4.5 Å. The trajectory is sampled every 100 ps for calculating the RMSD values shown here.

References

1. Grycuk, T., 2003. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. J. Chem. Phys. 119:4817–4826.