JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Efficient Conformational Sampling in Explicit Solvent Using a Hybrid Replica Exchange Molecular Dynamics Method

Sidhartha Chaudhury,*[,†] Mark A. Olson,[‡] Gregory Tawa,[†] Anders Wallqvist,[†] and Michael S. Lee[§]

[†]Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S Army Medical Research and Materiel Command, Fort Detrick, Maryland

[‡]Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland

[§]Computational Sciences and Engineering Branch, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland

Supporting Information

ABSTRACT: Temperature-based replica-exchange molecular dynamics (REMD), in which multiple simultaneous simulations, or replicas, are run at a range of temperatures, has become increasingly popular for exploring the energy landscape of biomolecular systems. The practical application of REMD toward systems of biomedical interest is often limited by the rapidly increasing number of replicas needed to model systems of larger size. Continuum solvent models, which replace the explicit modeling of solvent molecules with a mean-field approximation of solvation, decrease system size and correspondingly, the number of replicas, but can sometimes produce distortions of the free energy landscape. We present a hybrid implicit/explicit solvent REMD method in CHARMM in which replicas run in a purely explicit solvent regime while exchanges are implemented with a high-density GBMV2 implicit solvation model. Such a hybrid approach may be able to decrease the number of replicas needed to model larger systems while maintaining the accuracy of explicit solvent simulations. Toward that end, we run REMD using implicit solvent, explicit solvent, and our hybrid method, on three model systems: alanine dipeptide, a zwitterionic tetra-peptide, and a 10-residue β-hairpin peptide. We compare free energy landscape in each system derived from a variety of metrics including dihedral torsion angles, salt-bridge distance, and folding stability, and perform clustering to characterize the resulting structural ensembles. Our results identify discrepancies in the free-energy landscape between implicit and explicit solvent and evaluate the capability of the hybrid approach to decrease the number of replicas needed for REMD while reproducing the energy landscape of explicit solvent simulations.

INTRODUCTION

Molecular dynamics (MD) simulations of proteins can provide deep insight into the microscopic mechanisms that guide their structure and function. Accurate modeling of such systems typically requires an atomistic representation of both the protein molecule and thousands of individual solvent molecules that make up its local aqueous environment. The energy landscape of these complex systems contain innumerable local minima which hinders conformational sampling and present a major challenge to accurate and efficient modeling. A number of enhanced sampling methods have been developed to address this issue. These include replica exchange,[1] accelerated molecular dynamics,[2] self-guided molecular dynamics,[3] potential smoothing,[4] locally enhanced sampling,[5] and continuum solvent models.

Temperature-based replica-exchange molecular dynamics (REMD) has become an increasingly popular method for overcoming these challenges in protein modeling for a wide range of applications, from simulating equilibrium behavior to predicting structure.[6] REMD typically involves running multiple simultaneous simulations, or replicas, at a wide-range of temperatures, while allowing exchanges of temperatures between these replicas. The exchanges between two neighboring temperature replicas are attempted at regular intervals and made when the Metropolis criterion is met. This criterion relates the relative probability of finding each conformation at a given temperature as a function of their respective energies. REMD allows conformations to effectively percolate up and down the ladder of temperatures. Higher temperature replicas rapidly overcome potential energy barriers that impede sampling while lower temperature replicas sample lower-energy conformations that are relevant to a room-temperature ensemble.

Although traditional REMD has enjoyed success in modeling peptides and small proteins, as the system size grows larger, REMD becomes increasingly impractical. As larger proteins are modeled, larger solvation boxes with a greater number of solvent molecules are needed. This quickly increases the size and complexity of the system. The number of replicas required for efficient REMD is a function of the overlap in energy distribution between neighboring replicas. As the system size grows, the difference in energy between a given temperature range increases, and a larger number of replicas are needed to span that range.[7,8] An explicit solvent system for even a small protein, such as the 20-residue Trp-cage mini-protein requires more than 40 replicas;[9] for pharmacologically relevant proteins which are often several hundred residues in length, potentially hundreds of replicas may be required. Alternative replica-exchange methods have been developed that decrease the necessary number of replicas including nonequilibrium replica exchange[10] and biasing-potential replica exchange,[11] they are generally incompatible with

Received: July 29, 2011
Published: December 01, 2011

traditional temperature-based replica exchange of protein-sized systems.

In explicit solvent systems, the energy is dominated by bulk solvent−solvent interactions. One common approach to making these systems more tractable for simulation is by replacing the solvent molecules with continuum solvation model that approximates the free energy associated with solvent−solvent and solvent−protein interactions. The generalized-Born (GB) implicit solvent model is an accurate, efficient, approximation of the more rigorous but computationally expensive Poisson−Boltzmann (PB) continuum electrostatic model.[12] In a typical GBSA (Generalized-Born Surface Area) implicit solvation model, the polar contribution to the solvation free energy is calculated using the GB approximation, while the nonpolar contribution, also termed the "hydrophobic" effect, is captured by a surface area term. The decreased system size due to the absence of solvent molecules significantly reduces complexity. For example, a 10-residue peptide, which requires 24 replicas in explicit solvent,[13] requires only 8 replicas using a continuum solvation model.[14,15] Trp-cage, which requires over 40 replicas in explicit solvent,[9] required 17 replicas using the GB implicit solvation model.[16]

Although implicit solvent models allow for significantly faster simulations, there are a number of reported discrepancies between results from implicit solvent simulations and the explicit solvent simulations they were intended to reproduce. For a number of force fields including CHARMM22 and Amber ff09, simulations of alanine dipeptide show that GB solvation disproportionately favors α-helical dihedral conformations over β or PPII conformations.[17−19] Other studies have reported that GB models can induce overstabilized salt-bridges.[20,21] Simulations of short peptides have revealed discrepancies in ensemble properties such as end-to-end distance and radius of gyration (Rg) distributions.[17,22] Zhou and Berne showed significant differences in the free energy landscape of the C-terminal β-hairpin of protein G derived from an implicit GB model and explicit solvent.[23] Finally, Lee and Olson shows that the Trp-cage mini-protein has three distinct stable conformations in GB[16] which is not observed in explicit solvent.[9]

The discrepancies in simulation behavior between implicit and explicit solvent models can be attributed to two related issues. First, do continuum solvation models, in which there is a discrete dielectric boundary separating protein and solvent volumes, accurately reflect the solvation free energies for a given protein conformation? Second, does the surface area term in the GBSA model capture the complex effects of explicit solvent molecules on the conformational space sampled by the protein in an implicit solvent simulation? We hypothesize that differences in conformational sampling are primarily responsible for the discrepancies in implicit solvent simulations and that sampling in explicit solvent should be sufficient to reproduce explicit solvent simulation results.

Toward that end, we implement a hybrid implicit/explicit solvation model for REMD in CHARMM.[24] In this method, each replica is run exclusively in explicit solvent using the CHARMM22 force field. During an exchange attempt, the solvent molecules are removed, and the energy of the reduced system is calculated using the CHARMM22 force field with the GBMV2 implicit solvation model.[12] Once an exchange has been completed, the solvent molecules are reinserted into the system, at their previous positions and velocities, and the replica resumes its course. Since each replica is run in explicit solvent in its entirety, sampling is done exclusively along the free energy landscape of the explicit solvent system. Since exchanges are calculated using an implicit solvation representation of the system, the energy distribution of each replica is roughly similar to the distribution in the implicit solvent system. This reduces the number replicas needed to span a given temperature range. Theoretically, the GB solvation free energy calculation should approximate the free energy contribution of all solvent molecules over all solvent molecule degrees of freedom. Ultimately, by sampling exclusively in the explicit solvent domain while calculating energies for exchanges using an implicit solvation model, we aim to reduce the number of replicas needed for efficient REMD while accurately reproducing the results of the more computational expensive explicit REMD simulation.

We apply our method to three model systems to measure its efficiency and effectiveness in reproducing results derived from explicit REMD simulations. These systems include: blocked alanine dipeptide (Ace-A-Nme), zwitterionic tetrapeptide Ace-RAAE-Nme), and chignolin, a 10-residue β-hairpin peptide (GYDPETGTWG). All of these systems have been modeled in previous MD simulation studies and each has shown different behavior in implicit and explicit solvation models.[13,17−20] For each system, we carry out a fully implicit REMD, a hybrid REMD, and a fully explicit REMD simulation and calculate the resulting energy landscape in terms of dihedral torsion angle, salt-bridge stability, and peptide folding stability. Finally, we identify discrepancies in the energy landscapes from the three solvent models and evaluate the ability of the hybrid method in reproducing explicit solvent results at a fraction of the computational cost.

There have been two prior implementations of hybrid implicit/explicit solvation in REMD. The first implementation, by Okur et al.,[25] used the GB implicit solvation model from Amber to calculate the solvation energy of a reduced representation of the system that included at least two shells of solvent molecules during exchanges. They modeled alanine polypeptides of various lengths and found that, for small systems, there were significant efficiency gains when using the hybrid approach. However, the necessity of including multiple layers of solvent molecules in their method limits the efficiency gains and introduces nontrivial technical challenges in the application to larger protein systems. In another implementation of the hybrid implicit/explicit solvation in REMD, Mu et al.[22] used a reparameterized version of the computationally expensive PB solvation model to circumvent the need to include multiple layers of solvent molecules during exchanges. They applied their method to a range of systems including small peptides and the Trp-cage mini-protein and showed results comparable with explicit REMD.[13,22,26] In our implementation of the hybrid REMD method, we apply the GBMV2[12] implicit solvation model to leverage the computational speed and accuracy of GB with the simplicity of excluding all solvent molecules in the energy calculation. Since GB, unlike PB, is compatible with implicit solvent simulations, we will be able to evaluate the accuracy of the hybrid method vs the GB method compared to an explicit solvent benchmark.

METHODS

Temperature-Based Replica Exchange. The replica-exchange protocol involves running a number of independent simulations, or replicas, at a range of temperatures that periodically exchange temperatures. Each replica, a, exchanges its temperature with another replica, b, if $\Delta_{ab} < 0$ or $\exp(-\Delta_{ab})$ is greater than a random number generated uniformly between

Table 1. Summary of REMD Simulations

system	method	temperature range (K)	length (ns)	no. of clients	N_{water}	box size (Å)	CPU-hr/ns
Ace-A-Nme	implicit	298− 500	10	4			10
	hybrid	298−500	10	4	693	27.53	94
	explicit	298−500	10	24	693	27.53	550
Ace-RAAE-Nme	implicit	298−500	40	5			17
	hybrid	298−500	40	5	1324	34.09	200
	explicit	298−500	40	24	1324	34.09	990
GYDPETGTGWG	implicit	250−450	100	7			39
	hybrid	250−450	100	7	1180	32.92	270
	explicit	250−450	100	32	1180	32.92	1200

0 and 1, where Δ_{ab} is defined as

$$\Delta_{ab} = (E_b - E_a)\left(\frac{1}{k_B T_a} - \frac{1}{k_B T_b}\right) \quad (1)$$

where E_a is the energy of replica a corresponding to temperature T_a and k_B is Boltzmann's constant. In practice, exchanges are only attempted between replicas in adjacent temperature windows.

Simulation Setup. We studied the following peptide systems in the present work: blocked alanine dipeptide (Ace-A-Nme), zwitterionic tetrapeptide Ace-RAAE-Nme), and chignolin, a 10-residue β-hairpin peptide (GYDPETGTWG). These systems were selected to compare the three REMD methods: implicit REMD (imp-REMD), hybrid REMD (hyb-REMD), and explicit REMD (exp-REMD), over a diverse range of ensemble measurements as well as in protein folding and structure prediction. Details of the MD methods used in this work are as follows.

We performed MD simulations with the CHARMM program,[24] version c35b3. We used the CHARMM22 force field with the CMAP backbone dihedral cross-term extension.[27] We set an integration time step of 2 fs and applied the SHAKE[28] algorithm to fix all covalent bonds with hydrogen atoms. Nonbonded electrostatics and van der Waals interactions were truncated smoothly from 12 to 14 Å. For implicit solvent simulations, we used Langevin dynamics with a friction coefficient, γ, set to 10 ps^{-1} for all heavy atoms. For implicit solvent simulations, we used the GBMV2 model with default parameters except for the MMTSB option *gbmvaig* which was set to 97, which greatly increases the grid resolution and consequently the accuracy of the GB calculation during exchange attempts in the hybrid simulation. A surface tension value of 0.00542 kcal/(mol-Å^2) was used for the solvent accessible surface area nonpolar solvation term. For hybrid and explicit solvent simulations we used a Nose-Hoover thermostat with a temperature coupling constant of 50 kcal/s^2.

Replica exchange was performed using a modification of the MMTSB toolset[29] script *aarex.pl*. Analysis of replica-exchange results was facilitated by a modified version of the MMTSB script *rexanalysis.pl*, which automates WHAM analysis of the REMD simulation data. The temperature ranges, simulation lengths, system sizes, and dimensions are listed in Table 1. The number of replicas used varied from system to system, depending on its size. For alanine dipeptide, zwitterionic tetrapeptide, pentapeptide, and β-hairpin peptide, the number of replicas four, five, five, and seven, for implicit solvent and hybrid simulations and 24, 24, 24, and 32, for explicit solvent simulations. In the hybrid and explicit solvent simulations, peptides were solvated in a cubic box with 693, 1324, 1103, and 1180 TIP3P water molecules respectively, and the minimal number of sodium or chloride counterions were added to neutralize the system. We ran simulations to lengths described in Table 1, and exchanges between replicas were attempted once every 1 ps. In implicit solvent simulations, all structures were equilibrated for 4 ps prior to the simulation. For the hybrid and explicit solvent simulations, we equilibrated the solvated systems by running 5 ps simulations at 50, 100, 150, 200, 250, 275, and 298 K, at a constant pressure of 1 atm. The average cubic box dimensions from the final equilibration run were used for the constant volume simulations during replica exchange. The starting structure for all peptides, except chignolin was an extended conformation, with $\varphi = -150°$ and $\psi = 150°$. The starting structure for chignolin was the first conformer in the NMR structure (PDB code 1UAO).[30]

Hybrid Replica Exchange. In traditional implicit or explicit solvent REMD, each replica is run in the implicit or explicit solvent regime, and the energy corresponding to each replica that is used to calculate exchange probability is derived from force field parameters used within each replica. In our hybrid replica exchange method, each replica runs identically to the explicit solvent REMD. During an exchange attempt however, all water molecules and counterions are removed from a given replica, and the energy of the replica is calculated using the force field parameters from implicit solvent. Once an exchange attempt is completed, all removed molecules are inserted back at the positions and velocities prior to the exchange attempt, and the replica continues on. The number of replicas needed for efficient REMD is a function of the distribution of energies that span the REMD temperature range, and smaller systems, such as those in implicit solvent, have a narrower range of energies over that temperature span. By sampling in explicit solvent, but calculating exchanges in implicit solvent, we are able to run an explicit solvent REMD simulation with the same number of replicas used in implicit solvent REMD. The energy calculation for explicit is shown in eq 2, and compared with the hybrid and implicit in eq 3:

$$E_{explicit} = U_{MM}^{prot-prot} + U_{MM}^{prot-solv} + U_{MM}^{solv-solv} \quad (2)$$

$$E_{implicit} = U_{MM}^{prot} + \Delta U_{GBSA}^{solv} \quad (3)$$

We implemented a "semi"-hybrid method based as outlined by Okur et al.[25] specifically to further investigate salt-bridge stabilization in the hybrid REMD method. The semihybrid method is identical to the hybrid method described above except that a prespecified number of water molecules closest to the protein are kept during the energy calculation in an exchange attempt. This semihybrid method was applied to the zwitterionic peptide system and the nearest 200 water molecules were kept during exchange calculations. MMTSB Perl scripts for both the hybrid

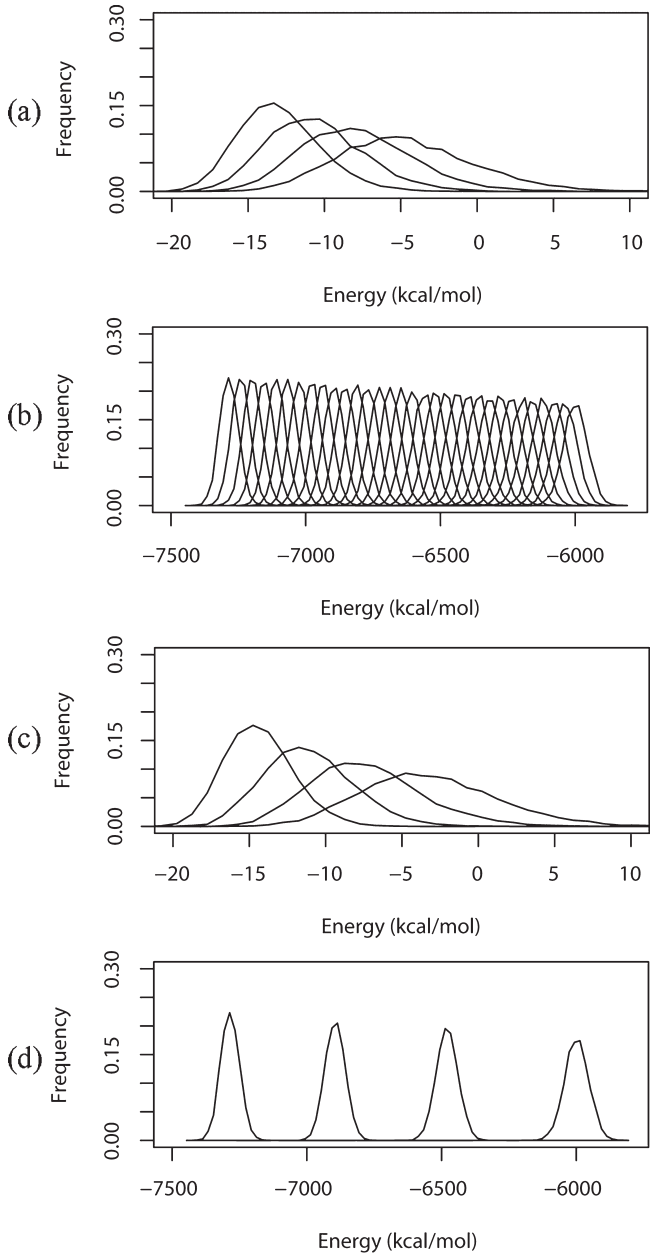

Figure 1. Energy distributions for each replica in (a) implicit, (b) explicit, and (c) hybrid methods, as well as (d) the explicit method with the same temperature replicas used in the implicit and hybrid methods. The distribution furthest to the left in each plot corresponds to the lowest temperature; the distribution furthest to the right corresponds to the highest temperature.

and semi-hybrid REMD methods are publicly available at our website: http://www.bhsai.org/downloads/hybrid_remd/.

Evaluation Metrics. We measured peptide structure and ensemble characteristics according to a variety of metrics. We described ensemble behavior largely through probability distributions of composite trajectories at 298 K and potential of mean force (PMF) calculations made using the temperature-based-weighted histogram analysis method[31] (WHAM) using all temperature replicas. All measurements and PMF calculations were facilitated by a modified version of the MMTSB toolset script *rexanalysis.pl*. The following structural measurements were binned in the WHAM algorithm. For alanine dipeptide, the φ and ψ backbone torsion angles were used. For the zwitterionic tetrapeptide, radius of gyration (R_g) and the salt-bridge distance (r_{sb}), defined as the distance between the C_ζ atom of Arg1 and C_δ atom of Glu4, were used. For chignolin, the C_α root mean squared distance (rmsd) to the first conformer in the NMR structure[30] (PDB code 1UAO), following optimal superposition, was used, as well as distance between the N atom of Gly1 and the C atom of Gly10. We calculated melting curves based on a WHAM analysis at 10 K intervals from 250 to 450 K. Standard deviations, shown in parentheses, are determined from running the identical analysis over the trajectory data blocked into thirds.

We pooled 250 ps snapshots from composite intermediate temperature trajectories of chignolin from all three REMD mehods and clustered the resulting peptides structures using the jclust hierarchical clustering algorithm using the MMTSB Tool Set script *cluster.pl*. We limited analysis to three levels of clustering. The top-level clustering identified two clusters, one corresponding to the folded state and the other to the unfolded state. The bottom-level cluster centers derived from the folded and unfolded clusters were then reclustered using a fixed-radius algorithm with a radius of 0.5 Å and 3.0 Å C_α rmsd for folded and unfolded clusters. Clusters representing less than 2 ns (8 snapshots) of simulation time for at least one method were thrown out of the analysis. Approximately 99% of snapshots were represented by one of the bottom-level clusters. Cluster centers are represented by the snapshot in each cluster with the lowest C_α rmsd to that cluster's centroid coordinates.

RESULTS

Computational Efficiency. The primary goal of the hybrid REMD method is to reduce the number of replicas needed to run in explicit solvent while maintaining the overall accuracy of explicit solvent MD. The number of replicas needed for efficient exchanges in REMD is a function of the overlap in energy distributions of adjacent replicas in a given temperature range, the larger the overlap, the higher the exchange rate.[8] Figure 1 illustrates the energy distribution from the alanine dipeptide simulation for implicit solvent, hybrid, and explicit solvent REMD methods, as well as the explicit method only the temperature clients used in the implicit and hybrid methods. In the implicit solvent and hybrid REMD methods, the energies span from appoximately −20 to +10 kcal/mol between the 298 and 500 K temperatures. For the explicit solvent method, these energies span from −7400 to −5800 kcal/mol over that same temperature span. The slightly lower energy distribution of the hybrid method compared to the implicit method can be attributed to the use of a Nose-Hoover thermostat for a solvated system consisting of 2098 atoms in hybrid compared to a Langevin thermostat for an unsolvated system of 22 atoms in implicit solvent.

The results show that the hybrid method is successful in reproducing the energy distributions of the implicit solvent method and the corresponding exchange rates, while running each replica in explicit solvent. It is clear that running four replicas in explicit REMD is insufficient for adequate exchanges as there is a large gap between the energy distributions of adjacent temperature replicas. The results suggest that the hybrid method can run explicit solvent replicas at approximately 1/5th of the computational cost of the explicit method for this system. These results are representative of all the systems modeled in this study.

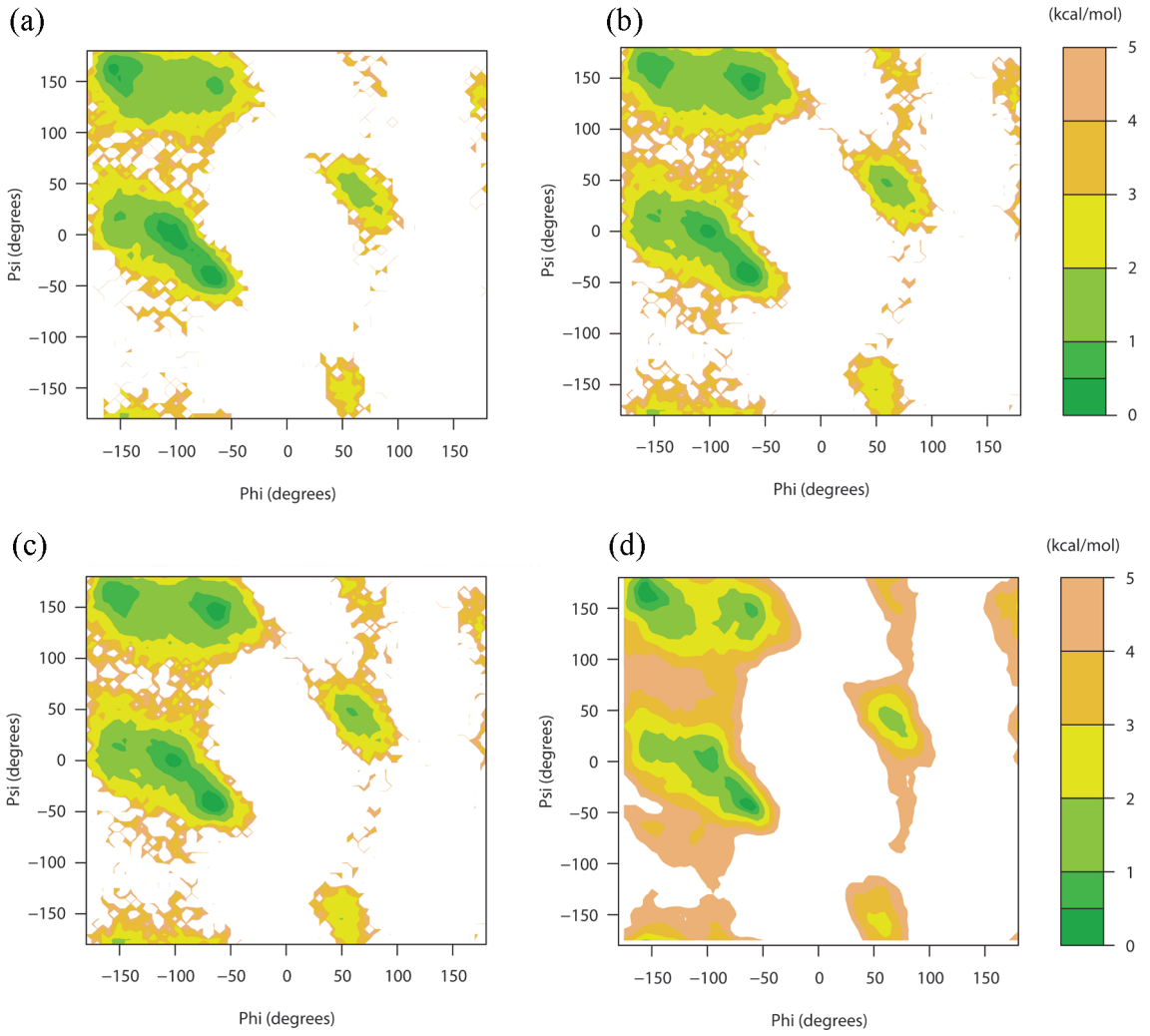

Figure 2. PMF with regard to backbone torsion angles φ and ψ for (a) implicit, (b) explicit, and (c) hybrid REMD methods. For comparison, the (d) adiabatic map generated using the same force field parameters as in implicit method for all values of φ and ψ, in 5° bins, confirms that all major minima are sampled.

Alanine Dipeptide Simulation. Alanine dipeptide has long been used as a model system for studying protein backbone conformational sampling in molecular dynamics.[19] A number of studies have highlighted small but significant differences in conformational sampling based on the solvent model used. We ran 10 ns simulations starting from the extended peptide conformation using the implicit, hybrid, and explicit REMD methods and collected data from 2 to 10 ns. From each of the three methods, we computed the resulting PMF along the φ and ψ backbone torsion angles to identify any systematic differences in the conformational free energy landscape between implicit and explicit methods and evaluate the ability of the hybrid method to accommodate these differences in accurately reproducing the results of the explicit solvent simulation. The three PMF diagrams are illustrated in Figure 2.

Overall, the free energy landscapes are similar for all three methods with the right-handed α-helical region α_R (−65°, −50°) as the most favorable, but the results confirm small but significant differences between the implicit solvent and explicit solvent energy landscape. Specifically, in implicit solvent, the PPII region (−65°, 150°) has lower energy relative to the β-sheet region (−150°, 150°), and α-helical region α_1 (0°, 100°) has lower energy compared to explicit solvent. Finally, the left-handed α-helical region α_L is higher energy in the implicit solvent. The hybrid method, despite using the implicit solvent model during exchanges, successfully reproduces the explicit solvent energy landscape, and matches the explicit solvent results with regard to these discrepancies.

Table 2. Distribution of Conformations at 298 K for the Three REMD Methods for Alanine Dipeptide

	conformation (%) [s.d.]		
method	α_1/α_L	β/PPII	α_R
imp-REMD	45.3 [3.0]	32.8 [3.2]	2.1 [0.5]
hyb-REMD	34.6 [2.5]	33.8 [1.1]	5.7 [1.9]
exp-REMD	36.4 [1.6]	34.8 [0.5]	4.5 [1.0]

The overall effect of the differences in the energy landscape of alanine dipeptide is an over-representation of α-helical conformations in the 298 K ensemble. Table 2 shows the conformation distribution between three largest conformational clusters, left-handed α-helical (α_1 and α_L basins), β-sheet

dx.doi.org/10.1021/ct200529b |J. Chem. Theory Comput. 2012, 8, 677–687

(β and PPII basins), and right-handed α-helical (α_R basin), and indicates that the ensemble at 298 K generated by the hybrid method most closely matches the explicit solvent results. The results of the alanine dipeptide simulations demonstrate that there are systematic differences in conformational sampling between the implicit and explicit solvent models and that the explicit solvent sampling in the hybrid approach is able to successfully reproduce the overall free energy landscape of explicit method, despite using an implicit solvent model for energy calculation during exchanges.

Zwitterionic Tetra-Peptide Simulation. A number of studies have identified the overstabilization of salt-bridges as a significant discrepancy between implicit and explicit solvent simulations. We used the three REMD methods with a zwitterionic tetrapeptide that is flanked by a positively charged arginine residue and a negatively charged glutamate residue. We ran 50 ns simulations starting from an extended peptide conformation for each the three REMD methods and collected data from 10 to 50 ns. We calculated the PMF for each simulation along the salt-bridge distance, defined as the distance between C_ζ atom of Arg 1 and C_δ atom of Glu 4 (Figure 3) and compare overall percentage of conformations that contain a salt-bridge in the composite 298 K trajectory for each method in Table 2.

Figure 3 illustrates the PMF with respect to salt-bridge distance for all three methods, normalized to the height of the energy barrier at 5.0 Å. A deep energy well is observed at a distance of approximately 3.5 to 4.0 Å, corresponding to the formation of the salt-bridge. The relative depths of this minimum is −2.0, −3.1, and −3.2 kcal/mol for explicit, hybrid, and implicit methods, respectively, indicating that both the hybrid and implicit REMD methods display substantial salt-bridge stabilization relative to the explicit REMD simulation. At salt-bridge distances greater than 4.0 Å, both the explicit and hybrid methods show two small minima at 6.6 and 9.5 Å corresponding to single and double-shelled solvent separation. Figure 4 illustrates representative structures from the 300 K composite trajectory from the hybrid method for the salt-bridge minimum, as well as at each of the two solvent separated minima. These minima highlight the structured water effects of explicitly sampled solvent in the hybrid approach.

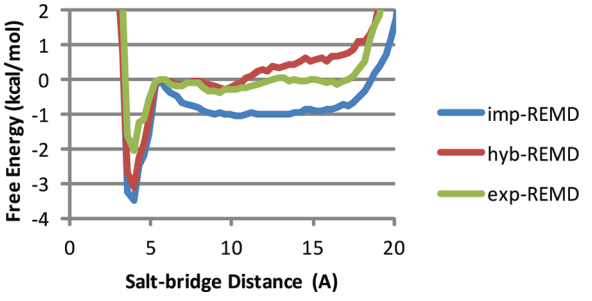

Figure 3. PMF at 298 K with respect to salt-bridge distance for the implicit, hybrid, and explicit REMD methods.

We compared the observed probability distribution of salt-bridge formation across the three REMD methods in Table 3. As expected, the explicit method demonstrated the lowest propensity for salt-bridge formation, at 58%. The implicit method showed substantially higher salt-bridge propensity at 80%, and surprisingly, the hybrid method showed the highest propensity for salt-bridge formation at 90%. This trend is reflected in the earlier PMFs which show that the hybrid method has, by far, the largest difference between the free energy of the salt-bridged conformations and the rest of the ensemble.

Overall, the zwitterionic tetrapeptide results suggest that while the hybrid method is able to reproduce features of the energy landscape that result from explicit water modeling, such as solvent-separated minima, the resulting thermodynamics can be profoundly affected by the implicit GB model used in the hybrid scheme. We implemented a 'semi' hybrid approach outlined by Okur et al.[25] specifically to study the role of that a shell of explicit water molecules can play on the salt-bridge stabilization observed in the hybrid. We carried out simulation with the same length, system-size and number of replicas as in the hybrid method and calculate PMFs with respect to both salt-bridge distance (Supporting Information Figure 1) and radius of gyration (Supporting Information Figure 2). The results show that the inclusion of an explicit water shell in the semihybrid method largely eliminated the salt-bridge stabilization observed in the hybrid method.

Chignolin β-Hairpin Peptide Simulation. The computationally designed 10-residue chignolin peptide forms a structurally defined β-hairpin and has been increasingly used to study peptide folding and thermodynamics in computational simulations. This system contains a number of protein-like features, including an antiparallel β-sheet with a β-turn, a salt-bridge, and a hydrophobic

Table 3. Distribution of Salt-Bridge Formation at 298 K for the Three REMD Methods for Zwitterionic Tetra-peptide

method	salt-bridge (%) [s.d.]
imp-REMD	80.1 [5.6]
hyb-REMD	90.1 [2.1]
exp-REMD	57.6 [0.0]

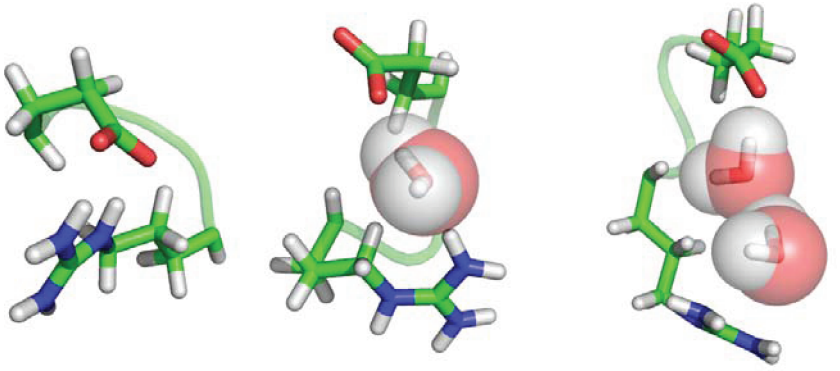

Figure 4. Representative structures from the hyb-REMD simulation at a salt-bridge distance of 3.8, 6.6, and 9.5 Å.

aromatic stacking interaction, making it an ideal test case for comparing solvent methods in REMD.

We ran 100 ns simulations, starting from the folded structure, for each of the three REMD methods. While all clients started from the folded conformation, by 50 ns each method had reached an equilibrium between the folded and unfolded states and all the analysis was done from data obtained from 50 to 100 ns. We calculated PMFs with respect to the Cα rmsd to the folded structure, the salt-bridge distance, and the minimum distance between the indole and phenol rings of Trp2 and Tyr8. We also extracted snapshots at 250 ps intervals and carried out hierarchical clustering to identify the major conformations populated by each simulation method at low (250 K), intermediate (370 K), and high temperatures (450 K). A summary of the results for chignolin can be found in Table 4.

Figure 5 illustrates the two-dimensional PMF with respect to salt-bridge distance and Cα rmsd for each of the three methods at 298 K. These PMFs can serve as a "fingerprint" that describes the energy landscape of the overall conformational ensemble for each method. All three PMFs show a deep basin corresponding with the folded structure with an rmsd of between 0.5 and 2.5 Å. Likewise, all three methods show a broad, shallow, basin corresponding to the unfolded state with rmsd great than 3.0 Å. The explicit method has significantly lower free energies in the unfolded basin compared to implicit method, with the hybrid method falling in between. Finally, both the explicit and hybrid methods show alternate minima corresponding to folded structures which lack a salt-bridge, and semifolded structures, with RMSDs between 2.5 and 3.5 Å that are largely absent in implicit simulation.

Table 4. Summary of Structural and Thermodynamic Properties for Chignolin

method	folded (%) [s.d.]	salt-bridge (%) [s.d.]	ΔG at 298 K (kcal/mol) [s.d.]	T_m (K)
imp-REMD	99.7 [0.2]	97.8 [1.1]	−2.8 [0.3]	375
hyb-REMD	94.0 [1.4]	93.7 [3.2]	−1.6 [0.1]	345
exp-REMD	75.0 [4.8]	90.3 [3.5]	−0.6 [0.1]	384

Using WHAM analysis, we calculated a theoretical melting curve (Figure 6) by generating a PMF with respect to rmsd at 10 K intervals from 250 to 450 K. We defined a folded conformation as one with a C_α rmsd of less than 2.5 Å, and calculated the fraction folded based on the free energy profiles at each temperature. The explicit method shows an atypical melting curve that is broad and relatively flat, with a melting temperature (T_m) of 384 K. This is in sharp contrast to the experimentally determined melting curve[30] but consistent with previous explicit solvent REMD simulations of chignolin.[13] The hybrid and implicit melting curves display a more typical sigmoidal form with T_m values of 375 and 345 K, respectively. Surprisingly, while the hybrid method's melting curve is a similar shape as the implicit method melting curve, it has a significantly lower T_m than both the implicit and explicit methods. The calculated folding free energy at 298 K was −0.6, −1.6, and −2.8 kcal/mol for explicit, hybrid, and implicit methods, respectively, with the hybrid method falling in between the implicit and explicit simulation results.

We calculated two-dimensional PMFs at the theoretical T_ms for each method (Figure 5) to observe the conformational landscape where the folded and unfolded populations are equal. The overall features of the energy landscape are largely similar to the PMF calculated at 298 K, but the relative depths of the energy basins differ. At the melting temperature, the hybrid method is strikingly similar to the explicit solvent method, with a substantial

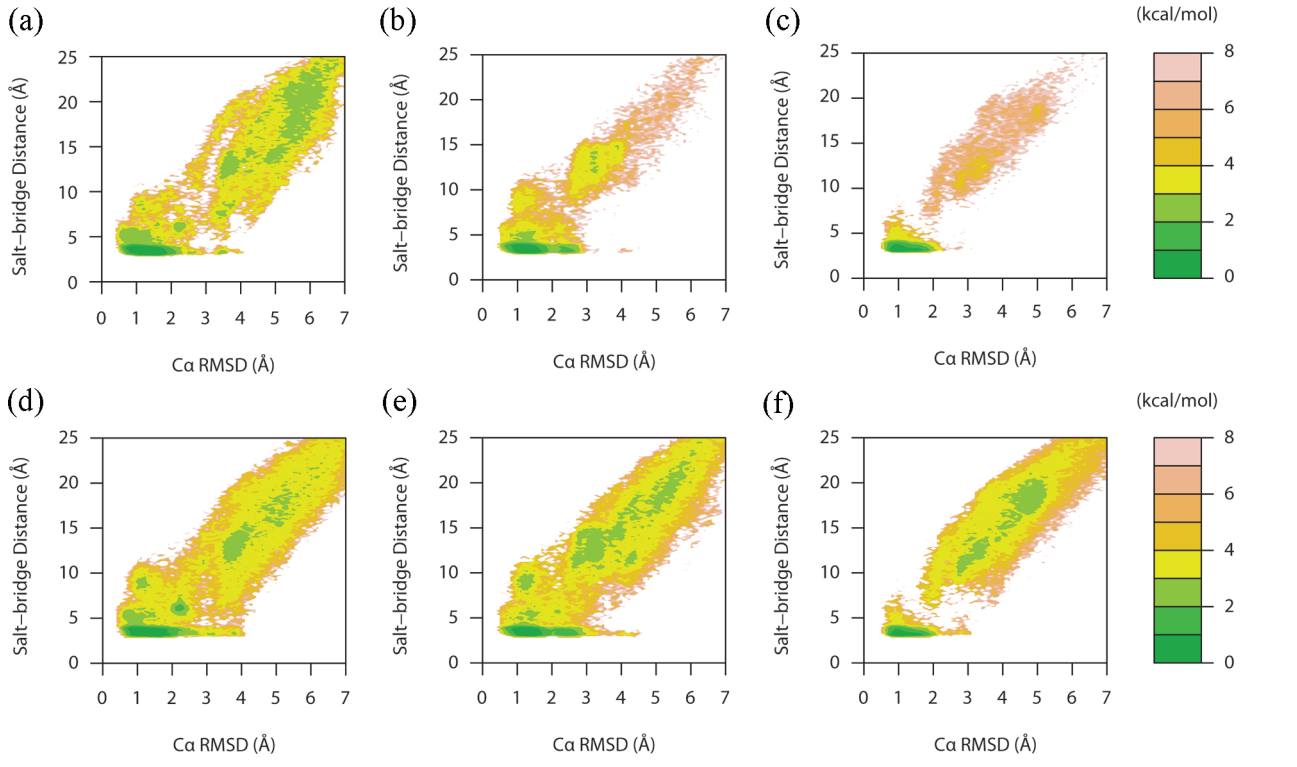

Figure 5. PMF of Cα rmsd and salt-bridge distance for explicit, hybrid, and implicit REMD methods at both 298 K (a, b, and c, respectively) and at each method's melting temperatures (d, e, and f, respectively).

dx.doi.org/10.1021/ct200529b |J. Chem. Theory Comput. 2012, 8, 677–687

population of semifolded states. By contrast, even at the melting temperature, the implicit method is almost devoid of these semifolded conformations, with a continuous high-energy barrier between the folded and unfolded states.

To understand this phenomenon, we identified the conformations that make up these folded, semifolded and unfolded populations by performing a clustering analysis. We selected trajectories for each simulation method from the temperature window that was closest to the respective T_m values, which were 369 K for the implicit and hybrid methods and 372 K for the explicit simulation, and collected snapshots at 250 ps intervals. We pooled these structures and carried out hierarchical clustering based on pairwise heavy-atom RMSDs. Figure 7 summarizes the clustering analysis, illustrating conformations at the center of each cluster, and their respective C_α rmsd to the folded structure. The top-level clustering resulted in a folded and unfolded cluster (hereafter referred to as parent clusters). The bottom-level cluster centers from these parent clusters were then reclustered using a fixed radius clustering algorithm (see Methods). Our analysis identified three folded clusters, F1, F2, and F3, whose cluster centers had native C_α RMSDs of 1.0 Å, 1.1 Å, and 1.4 Å, respectively. Satoh et al.[15] identified three hydrogen bonds that define the folded state: one between Asp3:O and Gly7:N, one between Asp3:N and Thr8:O, and one between the side-chain carboxyl group of Asp3 (Asp3:O_δ) and Glu5:N.

We identified four unfolded clusters, termed U1, U2, U3, and U4, with cluster centers that have RMSDs of 2.6, 4.2, 4.1, and 6.1 Å, respectively. Clusters U1 and U2 could be termed "semifolded" clusters as they contain native hydrogen bonds. 50% of conformations in cluster U1 contain the Asp3:O-Gly7:N hydrogen bond, and 16% contain both the Asp3:O-Gly7:N and Asp:N-Thr8:O

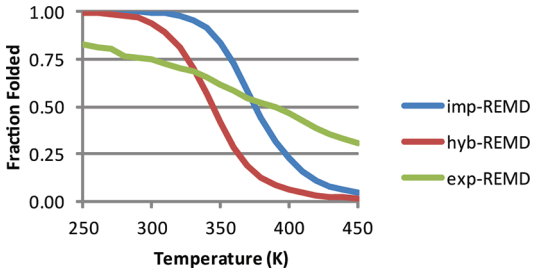

Figure 6. Simulated melting curves from the implicit, hybrid, and explicit REMD methods.

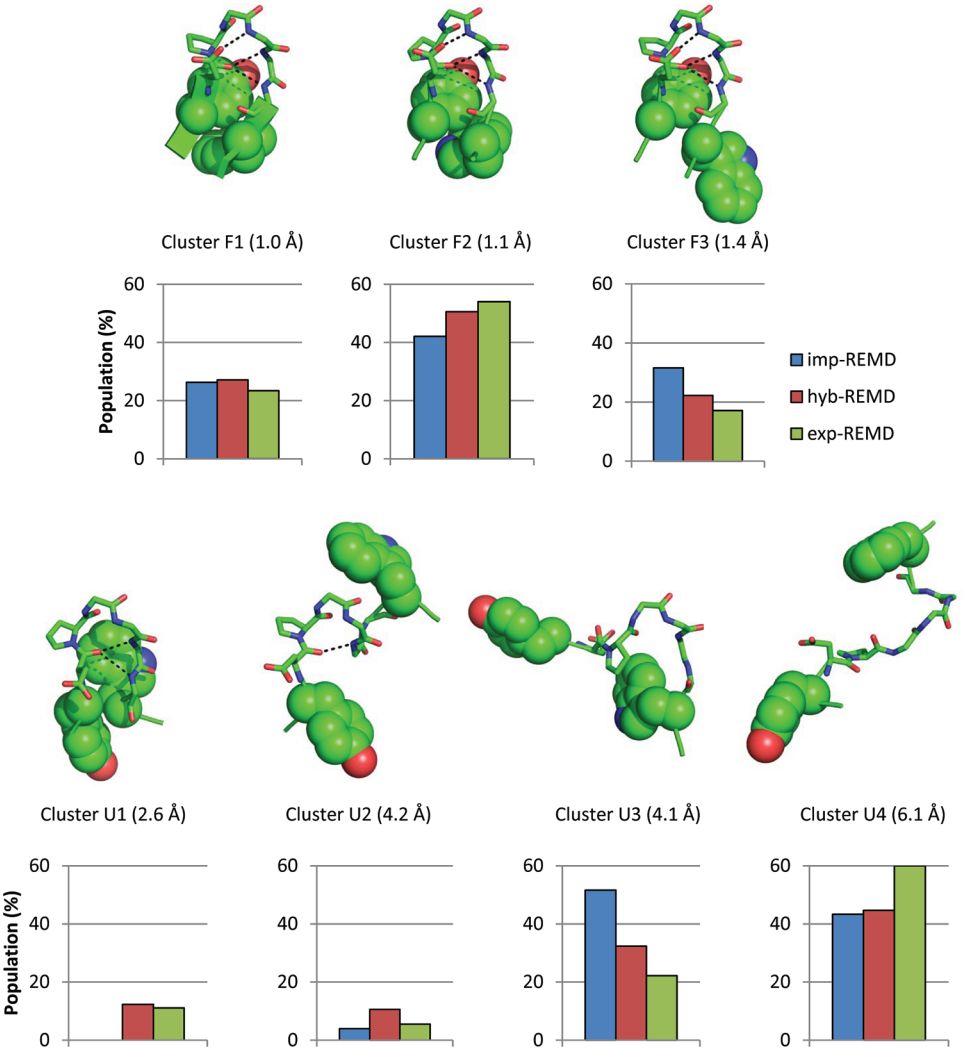

Figure 7. Structures of each cluster center, Cα rmsd to the folded structure, and relative population among folded (top) and unfolded (bottom) species for each method. Native hydrogen bonds are shown in black, aromatic residues Tyr2 and Trp9 are shown as spheres.

Table 5. Summary of Average Structural Properties and Hydrogen Bonding in Each Cluster for Chignolin

	average properties (S.D.)		native hydrogen bonds present (%)		
cluster	rmsd[a] (Å) [s.d.]	R_g (Å) [s.d.]	Asp3:O−Gly7:N	Asp3:N−Thr8:O	Asp3:O_δ−Glu5:N
F1	1.78 [0.30]	5.00 [0.09]	99	87	47
F2	2.10 [0.40]	5.00 [0.12]	93	68	50
F3	2.25 [0.53]	5.00 [0.15]	98	94	44
U1	4.18 [0.51]	5.33 [0.41]	50	16	39
U2	5.61 [0.52]	5.41 [0.44]	81	0	17
U3	5.51 [0.77]	6.22 [0.49]	0	0	0
U4	6.71 [0.83]	6.88 [0.84]	0	0	0

[a] Heavy-atom rmsd.

hydrogen bonds. Likewise, almost 80% of conformations in cluster U2 contain the Asp3:O-Gly7:N hydrogen bond. Clusters U3 and U4 lack any native hydrogen bonds, and conformations in cluster U4 were less compact than U3, with average radii of gyration of 6.9 Å (s.d. 0.8 Å) compared to 6.2 Å (s.d. 0.5 Å). Table 5 summarizes structural properties of conformations within each cluster.

In addition to clustering of the pooled snapshots of all three methods, we analyzed the composition of each cluster with respect to the method that each snapshot was derived from. Figure 7 shows the relative population of each subcluster with respect to parent cluster. Among folded parent cluster, all three methods show comparable composition within each cluster. Among the unfolded parent cluster, there are significant differences between the implicit method and the hybrid and explicit methods. In particular, 12% and 11% of the unfolded cluster members from hybrid and explicit methods fall into the semifolded U1 cluster, compared to 0% of the implicit. Likewise, 52% of unfolded cluster members from the implicit method fell into the unfolded cluster U3, compared to 32% and 22% from hybrid and explicit methods, respectively. These results suggest that the implicit method, significantly favors the unfolded state relative to the semifolded states, compared to the explicit and hybrid methods.

DISCUSSION

Continuum solvation models such as generalized Born have become increasingly popular as ways to make larger biomolecular systems tractable for computational modeling, particularly in REMD methods. However, there have been a number of reported differences between the free energy landscape of such implicit solvent models and the explicit solvent simulations they are designed to replicate. Here we present a hybrid REMD method that seeks to combine the computational efficiency of implicit solvent models while maintaining the accuracy of explicit solvent simulations. Toward that end, we model a number of peptide systems, identify discrepancies in the free energy landscape based on a number of metrics from backbone torsion angles, to salt-bridge distance, and folding stability, and evaluate the hybrid REMD method to reproduce explicit solvent simulation results.

As has been observed previously, the acceptance rate of REMD simulations is a function of the system size. Our hybrid REMD approach successfully reduces the number of clients necessary to run an explicit solvent simulation to that of an implicit solvent simulation and maintains the overall replica exchange acceptances rate of the implicit solvent model. This is achieved by reducing the system size associated with Metropolis exchanges to that of an implicit solvent system. Ultimately, the utility of the hybrid approach rests on whether this reduction in clients comes at the cost of decreased accuracy compared to explicit solvent simulations or experimental results.

In the smallest model system, alanine dipeptide, we observed small but significant differences in the energy landscape between implicit and explicit solvent simulations with regards to the φ and φ torsion angles. Overall, the implicit solvent model tended to overstabilize α-helical conformations relative to β-strand and PPII conformations, a result that has been observed previously.[18] Our hybrid approach successfully reproduced the energy landscape of the explicit solvent system with the relative energies at each minima, and population distributions within the margin of error compared to the explicit solvent simulations. Although this represents only a modest improvement in accuracy for this simple system, the cumulative sampling effects of such differences in larger, more complex biomolecular systems could lead to significant differences in the energy landscape.

One of the key reported differences between GB-based implicit solvent models and explicit solvent simulations is the overstabilization of salt-bridges. Using the zwitterionic tetra-peptide Ace-RAAE-Nme, we sought to identify the degree to which this overstabilization occurs in the GBMV2 implicit solvation model and evaluate the hybrid method's capabilities in overcoming this deficiency. As expected the implicit method showed a significant overstabilization of the salt-bridge and lacked the distinct solvent-separated minima that are present in the explicit method results. The hybrid method showed significant overstabilization of the salt-bridge, comparable to that of the implicit method, but was able to recover solvent-separated minima found in the explicit solvent PMF. These results suggest that while the hybrid method is able to recover features of the explicit solvent energy landscape, the GB solvent model used to calculate exchanges can still have a large effect on the resulting thermodynamics. Okur et al.[20] reported similar mixed results with their implementation of a hybrid solvation method, observing that a thorough analysis of salt-bridge overstabilization was confounded by systematic differences in backbone sampling in unrestrained implicit and explicit solvent REMD simulations.

We used the ten residue chignolin peptide as a model system because it contains protein-like structure features including an antiparallel β-sheet, a β-hairpin, a salt-bridge, and an hydrophobic packing interaction between two aromatic residues. We started each REMD simulation from the folded state and sought to reproduce explicit solvent structural and thermodynamic

properties of the system. Overall, the hybrid method reproduced a number of features of the explicit solvent energy landscape compared to the implicit solvent simulation, particularly with regards to semifolded conformations. Likewise, the calculated folding free energy from the hybrid method was closer to explicit solvent results and experiment, while the implicit solvent simulation resulted in a significantly overstabilized peptide. This could be due to the presence of a salt-bridge between the N and C-termini of the peptide in all three simulation methods which is largely unobserved in the NMR structure.[30]

A comparison of the simulated melting curves from each method for chignolin showed disparate results; the hybrid and implicit methods showed a sigmoidal curve, while the explicit method showed a flat curve, relatively insensitive to temperature. The unusual explicit solvent melting curve behavior, which has been observed previously, diverges significantly from the experimental melting curve[30] and suggests potential deficiencies in TIP3P water as a model for high-temperature water molecules. It is important to note that while explicit solvent REMD inherently accounts for some of the differences in the free energy of solvation at higher temperatures, the solvation free energies calculated using the GB-solvation model are temperature-invariant. As in the zwitterionic tetrapeptide system, the GB solvent model appears to have a significant effect on the thermodynamics of the hybrid system, influencing the path through temperature space that its explicit solvent trajectory travels, and ultimately affecting the composition and behavior of the room-temperature ensemble.

Finally, we sought to characterize the structural ensembles sampled through the three simulation methods for chignolin to see if any benefits arise from the explicit solvent sampling of the hybrid method versus implicit solvent. Specifically, we were interested in conformations that reflected transitions between the folded and unfolded states. We pooled snapshots from all three methods taken from composite trajectories at the respective melting temperatures and employed a hierarchical clustering strategy to characterize the resulting ensembles. We chose a two-stage clustering scheme that separately treated the folded and unfolded clusters and identified several folded and unfolded cluster centers. All three methods accessed the folded clusters with comparable frequency. Among the unfolded clusters, we identified two semifolded clusters that contained partial features of the folded peptide, including critical backbone hydrogen bonds that stabilize the β-hairpin. We found that the hybrid method reproduces the explicit solvent results in populating the semifolded cluster U1, while the implicit solvent fails to populate it to any significant degree. This semifolded cluster is highly similar to a semifolded structure observed by Mu et al. in an explicit REMD simulation of chignolin,[22] where it was deemed critical to the chignolin folding pathway. These results underscore the value of sampling in an explicit solvent regime in the hybrid method to reproduce ensembles generated in the computationally much more expensive explicit solvent simulations.

CONCLUSION

The hybrid approach succeeded in sampling explicit solvent simulations at a fraction of the computational cost. To the degree where implicit and explicit solvent simulations show different behavior, the hybrid method generally gives mixed results with respect to thermodynamic and structural properties of the room-temperature ensemble. However, its explicit modeling of solvent molecules recovers solvent-specific features of energy landscapes such as solvent-separated minima, and more importantly, appears to reproduce the structural sampling of explicit solvent simulations. In protein folding and structural refinement, implicit solvent simulations often fail to sample near-native conformations and explicit-solvent may provide folding pathways that implicit solvent cannot sample. The hybrid REMD approach provides a means for combining the enhanced sampling of REMD with explicit-solvent modeling at a fraction of the computational cost traditional methods, allowing for simulations with larger system sizes and longer time scales.

ASSOCIATED CONTENT

Supporting Information. Supporting Information includes two figures that illustrate the results of the semihybrid method on the salt-bridge stability of the zwitterionic tetra-peptide. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: schaudhury@bioanalysis.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Funding of this research was provided by the U.S. Department of Defense Threat Reduction Agency Grant TMTI0004_09_BH_T and the Department of Defense Biotechnology High Performance Computing Software Applications Institute. Computational time was provided by the U.S. Army Research Laboratory DoD Supercomputing Resource Center and the Maui High Performance Computing Center. Additional computational resources were provided by the Advanced Biomedical Computing Center. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense.

REFERENCES

(1) Ishikawa, Y.; Sugita, Y.; Nishikawa, T.; Okamoto, Y. Ab initio replica-exchange Monte Carlo method for cluster studies. *Chem. Phys. Lett.* **2001**, *33* (1−2), 199–206.

(2) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–29.

(3) Wu, X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512–518.

(4) Pappu, R. V.; Marshall, G. R.; Ponder, J. W. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nat. Struct. Biol.* **1999**, *6* (1), 50–5.

(5) Hornak, V.; Simmerling, C. Generation of accurate protein loop conformations through low-barrier molecular dynamics. *Proteins* **2003**, *51* (4), 577–90.

(6) Zhu, J.; Fan, H.; Periole, X.; Honig, B.; Mark, A. E. Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. *Proteins* **2008**, *72* (4), 1171–88.

(7) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. Exchange often and properly in replica exchange molecular dynamics. *J. Chem. Theory Comput.* **2010**, *6* (9), 2804–2808.

(8) Abraham, M. J.; Gready, J. E. Ensuring mixing efficiency of replica-exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **2008**, *4* (7), 1119–1128.

(9) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins* **2010**, *78* (8), 1889–99.

(10) Ballard, A. J.; Jarzynski, C. Replica exchange with nonequilibrium switches. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (30), 12224–9.

(11) Kannan, S.; Zacharias, M. Application of biasing-potential replica-exchange simulations for loop modeling and refinement of proteins in explicit solvent. *Proteins* **2010**, *78* (13), 2809–19.

(12) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., 3rd New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24* (11), 1348–56.

(13) Xu, W.; Lai, T.; Yang, Y.; Mu, Y. Reversible folding simulation by hybrid Hamiltonian replica exchange. *J. Chem. Phys.* **2008**, *128*, 175105–1−9.

(14) Suenaga, A.; Narumi, T.; Futatsugi, N.; Yanai, R.; Ohno, Y.; Okimoto, N.; Taiji, M. Folding dynamics of 10-residue beta-hairpin peptide chignolin. *Chem. Asian J.* **2007**, *2* (5), 591–8.

(15) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS Lett.* **2006**, *580* (14), 3422–6.

(16) Lee, M. S.; Olson, M. Protein folding simulations combining self-guided Langevin dynamics and temperature-based replica exchange. *J. Chem. Theory Comput.* **2010**, *6* (8), 2477–2487.

(17) Yeh, I. C.; Wallqvist, A. Structure and dynamics of end-to-end loop formation of the penta-peptide Cys-Ala-Gly-Gln-Trp in implicit solvents. *J. Phys. Chem. B* **2009**, *113* (36), 12382–90.

(18) Feig, M. Kinetics from implicit solvation simulations of biomolecules as a function of viscosity. *J. Chem. Theory Comput.* **2007**, *3*, 1734–1748.

(19) Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* **2008**, *4*, 1555–1564.

(20) Okur, A.; Wickstrom, L.; Simmerling, C. Evaluation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *J. Chem. Theory Comput.* **2008**, *4*, 488–498.

(21) Geney, R.; Layten, M.; Gomperts, R.; Hornak, V.; Simmerling, C. Investigation of salt bridge stability in a generalized Born solvent model. *J. Chem. Theory Comput.* **2006**, *2*, 115–127.

(22) Mu, Y.; Yang, Y.; Xu, W. Hybrid Hamiltonian replica exchange molecular dynamics simulation method employing the Poisson−Boltzmann model. *J. Chem. Phys.* **2007**, *127*, 084119–1−7.

(23) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12777–82.

(24) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545–614.

(25) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput.* **2006**, *2* (2), 420–433.

(26) Xu, W.; Mu, Y. Ab initio folding simulation of Trpcage by replica exchange with hybrid Hamiltonian. *Biophys. Chem.* **2008**, *137* (2−3), 116–25.

(27) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400–15.

(28) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(29) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **2004**, *22* (5), 377–95.

(30) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10-Residue folded peptide designed by segment statistics. *Structure* **2004**, *12* (8), 1507–18.

(31) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. Temperature weighted histogram analysis method, replica exchange, and transition paths. *J. Phys. Chem. B* **2005**, *109*, 6722–6731.