JCTC Journal of Chemical Theory and Computation

Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir

Asim Okur,[†] Daniel R. Roe,[†] Guanglei Cui,[†] Viktor Hornak,[#] and Carlos Simmerling*[,†,#,‡]

Department of Chemistry and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794, and Computational Science Center, Brookhaven National Laboratory, Upton, New York 11973

Received August 10, 2006

Abstract: Parallel tempering or replica-exchange molecular dynamics (REMD) significantly increases the efficiency of conformational sampling for complex molecular systems. However, obtaining converged data with REMD remains challenging, especially for large systems with complex topologies. We propose a new variant to REMD where the replicas are also permitted to exchange with an ensemble of structures that have been generated in advance using high-temperature MD simulations, similar in spirit to J-walking methods. We tested this approach on two non-trivial model systems, a β-hairpin and a 3-stranded β-sheet and compared the results to those obtained from very long (>100 ns) standard REMD simulations. The resulting ensembles were indistinguishable, including relative populations of different conformations on the unfolded state. The use of the reservoir is shown to significantly reduce the time required for convergence.

Introduction

Conformational sampling remains one of the largest challenges in simulating biologically relevant events in atomic detail. Even when a sufficiently accurate Hamiltonian of the system is used, the rugged and complex potential-energy surfaces usually result in simulations being trapped, prohibiting complete exploration of conformational space. Thus, significant effort has been put into devising efficient simulation strategies that locate low-energy minima for these complex systems. The challenges of conformational sampling have been discussed in several reviews.[1,2]

One major problem for molecular simulations is quasi-ergodicity where simulations may appear converged when observing some simulation parameters, but in reality large energy barriers may prevent them from sampling important regions of the energy landscape. Another simulation initiated in a different conformation may look converged as well, but comparison may show that only partial equilibration was achieved. An example of this behavior has been demonstrated by Smith et al. who reported that MD simulations of short peptides starting from different initial conformations were in poor agreement despite apparent convergence in some measured properties.[3]

One popular approach to overcoming quasi-ergodicity in a biomolecular simulation is the replica-exchange method .[4−8] In replica-exchange molecular dynamics (REMD)[9] (also known as parallel tempering[6]), a series of molecular dynamics simulations (replicas) are performed for the system of interest. In the original form of REMD, each replica is an independent realization of the system, coupled to a thermostat at a different temperature. The temperatures of the replicas span a range from low values of interest (experimentally accessible temperatures such as 280 or 300 K) up to high values (such as 600 K) at which the system is expected to rapidly overcome potential energy barriers that would otherwise impede conformational transitions on a computationally affordable time scale.

* To whom correspondence should be addressed. E-mail: carlos.simmerling@stonybrook.edu.
[†] Department of Chemistry, Stony Brook University.
[#] Center for Structural Biology, Stony Brook University.
[‡] Computational Science Center, Brookhaven National Laboratory.

10.1021/ct600263e CCC: $37.00
Published on Web 01/10/2007

At intervals during the otherwise standard simulations, conformations of the system being sampled at different temperatures are exchanged based on a Metropolis-type criterion[10] that considers the probability of sampling each conformation at the alternate temperature (further details are discussed in Methods). In this manner, REMD is hampered to a lesser degree by the local-minima problem, since simulations at low temperatures can escape kinetic traps by "jumping" directly to alternate minima being sampled at higher temperatures. Recently, it was shown that temperature distributions can be optimized to maximize the rate at which replicas traverse the temperature span. Moreover, the transition probability is constructed such that the canonical ensemble properties are maintained during each simulation, thus providing potentially useful information about conformational probabilities as a function of temperature. Because of these advantages, REMD has been widely applied to studies of peptide and small protein folding.[6,9,11−19]

For large systems, REMD can become intractable since the number of replicas needed to span a given temperature range increases with the square root of the number of degrees of freedom in the system.[20−23] Since the number of accessible conformations also typically increases with system size, the current computational cost for REMD simulations of large systems limits the simulation lengths to tens of nanoseconds per replica, which limits the ability to obtain converged ensembles for large systems. Several promising techniques have been proposed[20,24−27] to deal with this apparent disadvantage of REMD. To our knowledge, converged REMD simulations in explicit solvent from independent starting conformations have been reported only for short helical or unstructured peptides.[27−29]

Several studies have compared the sampling efficiencies of standard MD and REMD. Sanbonmatsu and Garcia reported a five-fold increase in sampled conformations using REMD over MD in the five-residue Met-enkephalin peptide in explicit solvent.[30] Zhang et al. showed that REMD enhances sampling over conventional MD by 15−70 times at different temperatures for the 21-residue Fs peptide in continuum solvent.[31] A recent study by Zuckerman and Lyman investigated the sampling efficiency of REMD through consideration of the rate acceleration afforded by increased temperature.[32] For slower converging systems (such as β-hairpins or more complex topologies where folding time is on the order of microseconds), REMD simulations are typically initiated from the native conformation (see recent example by Zhang et al.[33]) where unfolding through high-temperature replicas is obtained and temperature-dependent properties are calculated from the resulting structures.

REMD simulations increase conformational sampling over standard MD simulations, but reliable results for nontrivial systems are still challenging to obtain. It is possible that REMD does not provide even greater efficiency gains for peptides and proteins because the temperature dependence of the folding rate tends to be more weakly temperature dependent than the unfolding rate, as has been shown experimentally[34−38] and computationally.[39,40] When starting from non-native conformations, high-temperature replicas give limited advantage for finding native states since more minima on the free-energy landscape become accessible at higher temperatures, further complicating the search. Furthermore, when a high-temperature REMD replica locates a favorable low-energy basin (such as the native structure), this conformation is exchanged to lower temperature, and the high-temperature replica needs to repeat the search process. Importantly, during the search by the high-temperature replicas, all replicas continue to be simulated. Thus a very large set of simulations, all of which are long enough for the high-temperature replicas to sample multiple folding events, can be required to achieve correct Boltzmann-weighted ensembles across the range of replicas. From another perspective, REMD drives the generation of correct equilibrium ensembles of structures by employing an exchange criterion that explicitly assumes that structures being considered for exchange have Boltzmann-weighted probability of being sampled (see Methods for details). However, this assumption is only true after the generalized ensemble has already reached convergence and is typically incorrect at the start of the REMD simulation. Thus until all temperatures sample an equilibrium ensemble, none of the temperatures would be expected to have correct distributions because of coupling of replicas through an incorrect exchange probability.

An approach to the reduction of quasi-ergodicity that is conceptually similar to REMD was reported by Frantz et al. for Monte Carlo (MC) simulations of atomic clusters.[41] In their approach, called jump-walking (or J-walking), they coupled one MC simulation to another at higher temperature. Somewhat analogously to REMD, the low-temperature simulation was used to sample local minima and provide thermodynamic ensemble data at the temperature of interest, while the high-temperature simulation was used to facilitate barrier crossing. Periodically the low-temperature structures escape local minima by "jumping" to basins sampled at high temperatures. The Boltzmann distribution generated by the high-temperature walker becomes the sampling distribution for attempted jumps by the low-temperature walkers. One drawback is that too large a temperature difference results in poor acceptance probabilities for the jump, comparable to the need to optimize the spacing between REMD temperatures. Variations of the J-walking scheme were tested by employing high-temperature simulations on a different time scale than the low-temperature simulation or using multiple high-temperature simulations. They determined that the most efficient method is running the high-temperature walker to obtain an adequate distribution and using the stored conformations for jumps in a MC run at slightly lower temperature. The results of this lower-temperature run were then used as the seed set for a new J-walking run at even lower temperature. They validated this approach using simple double-well potentials where comparison to analytical results was possible and in simulations of Argon clusters of various sizes. Similar approaches to J-walking have been developed, such as smart walking (S-walking),[42] smart darting,[43] and cool walking .[44] The J-walking scheme has been adapted to REMD simulations that employ a resolution exchange scheme, where replicas were run using a coarse-grained

model to obtain conformations to be subsequently sampled by an all-atom model.[45,46]

Here we introduce a variant to REMD where we draw upon the strengths of the J-walking approach to overcome the slow convergence and high computational expense of REMD. Similar to J-walking, an ensemble of structures is generated using standard MD simulation at high temperature. Instead of the reduction of the temperature stepwise and re-equilibratation of the ensemble in stages, an REMD run is used to link, in a single step, the high-temperature ensemble to low temperatures of interest. Periodic exchanges are made between randomly chosen conformations from the reservoir set and the highest-temperature replica. This process formally provides correct ensembles at lower temperature with free energies that reflect the proper relative populations of minima. Importantly, the convergence speed of the REMD run is greatly enhanced since exchanges are attempted from an already converged Boltzmann ensemble and thus the exchange probabilities are correct at the start of the REMD run. We call this method reservoir REMD (R-REMD) because REMD is coupled to a high-temperature reservoir.

One major advantage of the reservoir approach with REMD is that a converged ensemble of conformations has to be generated only once and only for one temperature. After extensive conformational search at one temperature, the remaining temperatures can sample from and anneal these structures to rapidly construct equilibrium distributions consistent with their thermostat temperature. This is in contrast to the typical REMD approach where all replicas are run simultaneously, and the computational expense for running long simulations must be paid for each of the replicas even though only a few high-temperature ones may be contributing to the sampling of new basins. Another advantage is that the exchanges with the reservoir need not be time correlated with the replica simulations. Folding events sampled during reservoir generation can provide multiple native structures for the other replicas, in contrast to standard REMD where an independent folding event is required for each temperature that will have a substantial native population. Overall, the convergence rate for the set of replicas is greatly enhanced by exchanging with a previously converged ensemble.

We have implemented the reservoir REMD approach in the Amber[47] simulation package and have tested it on two models peptides, the trpzip2 β-hairpin[48] and the dPdP[49] three-stranded antiparallel β-sheet. These systems were selected because of the complexity and slow folding of β-sheets and hairpins as compared to that of α-helices, which fold rapidly enough that the performance advantage of R-REMD may not be apparent. For both systems, reservoir ensembles were generated at 400 K using the generalized Born[50] (GB) implicit solvent model using multiple simulations with different initial conditions. Subsequent R-REMD simulations were compared to standard REMD calculations with the same temperature ranges. In all cases, simulations were extended until close agreement was obtained between results obtained from independent runs with different initial structure ensembles (folded and unfolded). For both peptides, the use of reservoir structures is shown to provide the same structure ensembles and thermal melting profiles as standard REMD, with a reduction in overall computational cost of 5 to 20 times, including generation of the reservoir.

Methods

Replica Exchange Molecular Dynamics (REMD). We briefly summarize the key aspects of REMD as they relate to the present study. In standard parallel tempering or replica-exchange molecular dynamics,[6,9] the simulated system consists of M noninteracting copies (replicas) at M different temperatures. The positions, momenta and temperature for each replica are denoted by $q^{[i]}$, $p^{[i]}$, and T_m, $i = 1,..., M$ and $m = 1, ..., M$. The equilibrium probability for this generalized ensemble is

$$W(p^{[i]},q^{[i]},T_m) = \exp\left\{-\sum_{i=1}^{M}\frac{1}{k_B T_m}H(p^{[i]},q^{[i]})\right\} \quad (1)$$

where the Hamiltonian $H(p^{[i]},q^{[i]})$ is the sum of kinetic energy $K(p^{[i]})$ and potential energy $E(q^{[i]})$. For convenience, we denote $\{p^{[i]},q^{[i]}\}$ at temperature T_m with $x_m^{[i]}$ and further define $X = \{x_1^{[i(1)]}, ..., x_M^{[i(M)]}\}$ as one state of the generalized ensemble. We now consider exchanging a pair of replicas. Suppose we exchange replicas i and j, which are at temperatures T_m and T_n, respectively,

$$X = \{ ...; x_m^{[i]}; ...; x_n^{[j]}; ...\} \rightarrow X' = \{...; x_m^{[j]}; ...; x_n^{[i]}; ...\} \quad (2)$$

To maintain detailed balance of the generalized system, microscopic reversibility has to be satisfied, thus giving

$$W(X)\rho(X \rightarrow X') = W(X')\rho(X' \rightarrow X) \quad (3)$$

where $\rho(X \rightarrow X')$ is the exchange probability between two states X and X'.

A key step in the derivation of the exchange criterion[9] is the substitution of the Boltzmann factor for the weight of each conformation into eq 3, yielding eq 4. We note that this is not strictly correct until equilibrium has been reached, at which point the structures are actually considered for exchange with this probability.

$$\exp\left\{-\frac{1}{k_B T_m}H(p^{[i]}, q^{[i]}) - \frac{1}{k_B T_n}H(p^{[j]}, q^{[j]})\right\}\cdot\rho(X \rightarrow X') = \exp\left\{-\frac{1}{k_B T_m}H(p^{[j]}, q^{[j]}) - \frac{1}{k_B T_n}H(p^{[i]}, q^{[i]})\right\}\cdot\rho(X' \rightarrow X) \quad (4)$$

In the canonical ensemble, the potential energy, E, rather than the total Hamiltonian, H, can be used because the momentum can be integrated out .[9] By rearranging eq 4, one obtains the following Metropolis exchange probability

$$\rho = \min\left(1, \exp\left\{\left(\frac{1}{k_B T_m} - \frac{1}{k_B T_n}\right)(E(q^{[i]}) - E(q^{[j]}))\right\}\right) \quad (5)$$

It is important to reiterate that eq 4 is valid only for equilibrated ensembles that follow Boltzmann distributions. This assumption is true at the end of the simulation, and use of this exchange probability drives each replica toward adoption of the correct ensemble.

In standard REMD, several replicas at different temperatures are simulated simultaneously and independently for a chosen number of MD steps. Exchange between a pair of replicas is then attempted with a probability of success calculated from eq 5. If the exchange is accepted, the bath temperatures of these replicas will be swapped, and the velocities will be scaled accordingly. Otherwise, if the exchange is rejected, each replica will continue on its current trajectory with the same thermostat temperature.

Reservoir REMD (R-REMD). Reservoir REMD simulations (R-REMD) were run using same simulation parameters as standard REMD simulations. The only difference is that the highest temperature replica is replaced with a previously generated structure reservoir (replica R^N). Standard replicas (MD simulations) were used for each of the lower temperatures (replicas R^1 to R^{N-1}). Exchanges are attempted on the basis of the same criterion as used for standard REMD (eq 4). During the exchange attempts for replicas between R^1 and R^{N-1}, the exchange calculation is performed using current simulation coordinates. The only difference between R-REMD and REMD is when an exchange is attempted between replica R^{N-1} and the reservoir set R^N. The exchange attempt is made between the current structure of R^{N-1} and a randomly selected structure from the reservoir. If the exchange is accepted, the coordinates and velocities from R^N are sent to replica R^{N-1}. Formally the coordinates from replica R^{N-1} would be placed into the reservoir; however, for computational convenience, it was discarded since we assume that the reservoir constitutes a complete representation of the ensemble and that the inclusion of the new coordinates will have a negligible effect on the reservoir. Similarly, the chosen reservoir structure is left in the reservoir. Since the reservoir implemented for these tests is finite in size (10 000 structures), although the method is rigorous the present *implementation* does not formally obey detailed balance. For all practical purposes, the effect is minimal and such exchanges could readily be incorporated into the code.

Model Systems. The first model system chosen was the tryptophan zipper (trpzip) developed by Starovasnik and coworkers.[48] This β-hairpin structural motif is stabilized through cross-strand tryptophan pairs. Trpzip2 (SWTWENGKWTWK, with a type I′ β-turn at NG) has the most cooperative melting curve and highest stability (~90% at 300 K) among the trpzips and was selected for use in this study. Thermodynamic properties for this peptide have been determined by NMR and CD spectroscopy, and a family of structures was refined using restraints from NMR experiments[48] (PDB code 1LE1). The N-terminal of the peptide was acetylated and the C-terminal was amidated, in accordance with the experiments.

The second model system was created from the sequence of DPDP (VFITSdPGKTYTEVdPGOKILQ, dP = D-proline, O = ornithine), except that lysine was substituted for the ornithine. The replacement of ornithine with lysine in a related peptide analogous to the C-terminal hairpin of DPDP caused no detectable effect on the structure .[51] The termini were amidated and acetylated in accordance with experiments. DPDP was designed with a net charge of +2 to prevent aggregation, and our model retains this net charge.[49]

REMD Simulations. For both systems, standard REMD simulations were carried out with Amber, version 8.[47] For trpzip2, all covalent bonds were constrained using SHAKE .[52] For dPdP, only the bonds involving hydrogen atoms were constrained. A 2fs time step was used, and temperatures were maintained using weak coupling[53] to a bath with a time constant of 0.5 ps^{-1}. All nonbonded interactions were calculated at each time step (i.e., no cutoff was used). To permit comparison to our previously published data, both peptides were simulated with the Amber ff99 force field with modified backbone parameters .[54] Steepest-descent energy minimization was performed for both systems for 500 steps prior to REMD simulations. Both systems were simulated with the generalized Born solvation model[50] with GB^{HCT} [55] implementation in Amber. Scaling factors were taken from the TINKER modeling package .[56]

Standard REMD simulations were performed for both systems using 14 replicas for trpzip2 and 12 replicas for dPdP, covering a temperature range of ~260−570 K with an expected exchange probability of 15%. For trpzip2, additional replicas were manually placed between 300 and 370 K to increase statistics around the experimentally observed melting transition. Exchanges between neighboring replicas were attempted at 1ps intervals.

For both systems, two independent replica exchange simulations were run. For trpzip2, one simulation initiated all replicas in the published native conformation. The other simulation started from a compact non-native conformation where no hairpin backbone hydrogen bonds were present. Both REMD simulations were run to 155 000 exchange attempts (155 ns per replica). dPdP simulations were run as explained in our previous work,[17] with a simulation starting with all replicas in a fully extended structure and another with all replicas in a compact non-native structure. dPdP simulations were carried out for 170 000 exchange attempts (170 ns per replica).

Generation of Reservoir Structures. The reservoir structures were generated through molecular dynamics at 400 K with the same simulation parameters as used for REMD simulations. For trpzip2, four independent MD simulations each of ~38 ns in length were run starting from an extended conformation. Multiple folding and unfolding transitions were observed for each trajectory. For dPdP, a single long trajectory of 260 ns was generated. Multiple folding and unfolding transitions were observed. For both systems, velocities and coordinates were saved at 1 ps intervals. In the present implementation of R-REMD in Amber, coordinates and velocities for the reservoir were loaded into memory at the start of the R-REMD run. To minimize memory requirements, the reservoir ensembles were reduced to 10 000 structures by selecting equidistant snapshots from the trajectories.

Reservoir REMD Simulations. For trpzip2, four replicas were used below the 400 K reservoir with temperatures of 300, 323, 350, and 373 K. No additional replicas were used since these four replicas were sufficient to provide a 25−30% exchange ratio. Two sets of R-REMD simulations were

each run for 50 000 exchange attempts, starting from the same native or unfolded initial conformations as used for the standard REMD calculations.

Since dPdP is a larger system, the R-REMD simulations used 6 replicas below 400 K with the same temperature distribution as the standard REMD reported by Roe et al.[17] One R-REMD simulation starting from extended conformations was run for 50 000 exchange attempts.

Analysis. The trajectories obtained from standard and reservoir REMD simulations were analyzed using the Amber ptraj module. Trpzip2 simulations were compared to the experimentally determined native structure[48] (model 1 of PDB code 1LE1), where backbone rmsd's were calculated for residues 2−11. Terminal residues were omitted to remove the effects of fluctuations. An rmsd cutoff of 1.7 Å was used to determine native structures on the basis of the free-energy profile along rmsd where the native minimum reached up to 1.7 Å (data not shown). For dPdP, the fraction of native contacts were calculated, and the native population was calculated using a cutoff of 0.50 for both hairpin1 and hairpin2 contacts (as described in Roe et al.).[17]

Melting curves were generated by calculating the average population of native structures at each temperature. For trpzip2 simulations, data from the first 55 000 exchange attempts were discarded for each standard REMD simulation to remove initial structure bias. For dPdP, REMD simulations data from the first 20 000 exchange attempts were discarded.

Native fractions as a function of time were calculated by averaging the native population up to that time point for both systems using their respective criteria. For all systems, the rate of convergence was observed by comparing populations starting from different initial conformations. When both simulations show similar observables and a flat profile is obtained for all temperatures, the simulations are classified as converged.

Cluster analysis was performed as described previously[27] using the Moil-View program.[57] The trajectories from standard and reservoir REMD simulations were combined. Cluster analysis was performed on the combined set, and then normalized populations for each cluster type were calculated for each of the original simulations. This process permits direct comparison of the populations since the structure families are defined using the combined trajectories.

Results and Discussion

We apply the R-REMD method to two model systems (trpzip2 and dPdP) that we have studied previously using standard REMD. To validate the R-REMD approach, we first compare the resulting structure ensembles to those obtained with standard REMD to show that R-REMD provides accurate results. Next, we examine whether R-REMD provides these results more efficiently than standard REMD.

Trpzip2 REMD Simulations. We performed 2 independent REMD simulations of the trpzip2 peptide, one starting with all replicas in the published NMR structure (native) and one from a compact non-native structure. Both simulations were run ∼155 000 exchange attempts (equivalent to 155 ns per replica) where 14 replicas were used to cover a temperature range of 260−570K.

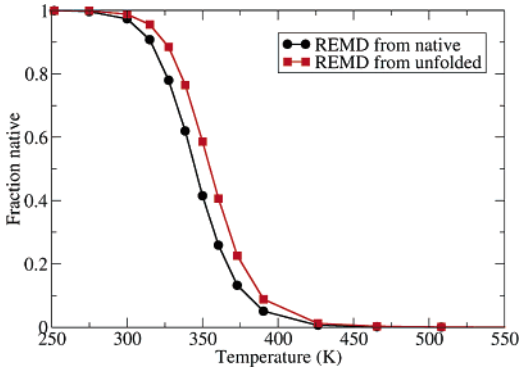

Figure 1. Melting curves for trpzip2 REMD simulations starting from native and unfolded conformations. Symbols represent temperatures at which simulation data is obtained. The similar profiles suggests that the data is reasonably well converged. Simulations show melting temperatures of 342.3 and 352.4 K, in excellent agreement with the experimentally measured value of 345 K.

Even though trpzip2 is a small system, long simulations were required to obtain good agreement between simulations with different initial conformations. Throughout the simulations the melting profiles were monitored and compared. After ∼150 ns, both REMD simulations showed similar melting profiles which no longer changed with increasing simulation times. The convergence rates of each simulation will be discussed later in this section. Since significant time was required to overcome the bias from initial conformations, data from the first 55 000 exchange attempts (55 ns) were discarded for constructing the melting curves (Figure 1). It should be noted that the amount discarded is larger than the total simulation time of most current published REMD studies. Simulations starting from unfolded conformations show slightly higher stability than those initiated with the native state, suggesting that these differences involve fluctuations in the data and do not reflect initial structure bias. As determined by fitting of the native fractions to the Gibbs−Helmholz equation, both simulations show comparable thermodynamic properties with melting temperatures of 342.4 and 352.4 K and ΔH_m of −15.90 and −16.46 kcal mol^{-1}. These values are in excellent agreement with the experimental melting temperature of 345 K and ΔH_m of −16.8 kcal mol^{-1}.[48] While the accuracy of the force field is not the subject of this study, it indicates that we are evaluating the performance of the R-REMD method under conditions that are relevant to experimental observations.

Testing the Accuracy of R-REMD. After the benchmark results were obtained using converged standard REMD simulations, we generated the high-temperature reservoir ensemble at 400 K. We chose 400 K because it is high enough to allow rapid conformational transitions, and it is well above the T_m, thus requiring R-REMD to significantly transform the reservoir ensemble to obtain accurate ensembles at lower temperatures.

Four standard molecular dynamics simulations were performed at 400 K using conditions identical to those for standard REMD simulations. Each simulation was run for ∼38 ns with a cumulative simulation time of ∼152 ns, where multiple folding and unfolding transitions were observed for

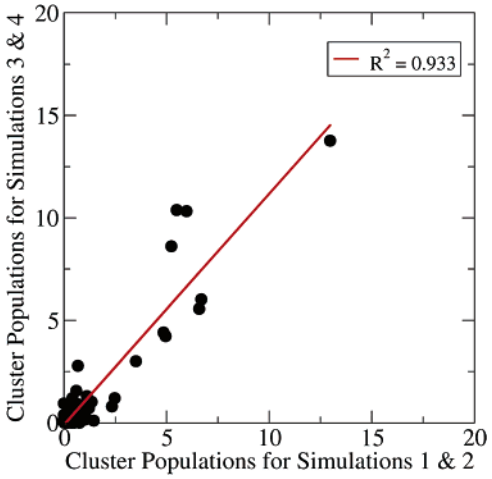

Figure 2. Populations of different trpzip2 structure clusters sampled by standard MD simulations. Populations of the first two trajectories are compared to populations of the same clusters in the remaining two trajectories. All clusters with large populations in runs 1 and 2 are also present with similar populations in runs 3 and 4, suggesting good convergence.

each trajectory (Figure S1). The presence of reversible folding transitions during standard MD is a reasonable indicator that the ensemble is fairly well converged (discussed in more detail below). Because of the elevated temperature, rapid unfolding takes place after each folding event, and the native population for each simulation is between 1 and 5%, in good agreement with the melting curves shown in Figure 1 (3.3 and 6.1% native populations at 400 K, calculated using the Gibbs−Helmholz equation and native fractions at the other temperatures).

Following our previously published work,[27] we evaluate the population of each cluster to determine whether independent simulations provide the same ensembles. This analysis is more rigorous than that comparing only native populations. Cluster analysis resulted in 136 structure families. In Figure 2, we compare the populations of each family sampled in the first two trajectories to the populations from the other two trajectories. A good correlation is observed, suggesting that the simulations not only sample the same types of structures, but that the relative population of each structure family is similar. While the composition of the unfolded ensemble will be discussed elsewhere, it is important to note that the most populated clusters (10−15% of the ensemble) are non-native at this elevated temperature, with a native population of only ∼3%.

This pool of 10 000 structures (coordinates and velocities) was used as the reservoir set for the R-REMD simulations. Four replicas were used with temperatures 300, 323, 350, and 373 K, where the 373 K replica periodically attempted to exchange with the 400 K reservoir as described in Methods. Two sets of R-REMD simulations with different initial structures were run for 50 000 exchange attempts. During the simulations, 25−30% exchange ratios are observed between replicas, and a 30% ratio is observed between the 373 K replica and the 400 K reservoir. Potential-energy overlaps were adequate for all temperature pairs (Figure S2).

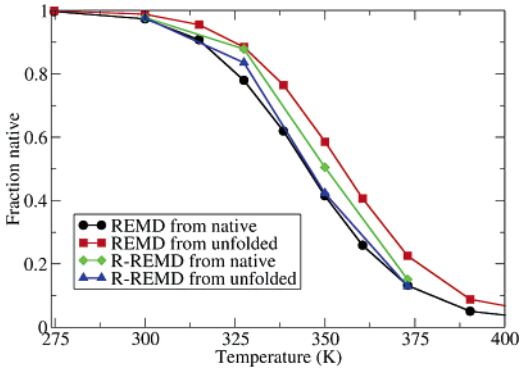

Figure 3. Thermal melting profiles for trpzip2 obtained from standard REMD (black and red) and R-REMD simulations (blue and green). Symbols represent temperatures at which simulation data is obtained. Standard REMD simulations are shown in black and red and R-REMD results are shown in green and blue. For easier comparison, only temperatures below 400 K are shown. Both R-REMD simulations are in good agreement with each other and lie fully within the precision range defined by the standard REMD results.

We next evaluated whether the use of the reservoir had any negative impact on the accuracy of the simulations. We calculated the thermal melting profiles for ensembles from the R-REMD simulations using the same procedure that was used for the standard REMD data. In Figure 3, we show the comparison of these melting curves to those from standard REMD. Excellent agreement is observed; the melting curves from the two R-REMD simulations lay within the bounds defined by the curves obtained from the two standard REMD simulations. Importantly, the R-REMD ensembles at low temperature are nearly fully native despite the low (3%) native population in the reservoir; thus, the REMD replicas are capable of accurately transforming the ensemble in the reservoir to what should be sampled at alternate temperatures. This result also suggests that it is possible to use this method for structure prediction, since the native conformation at low temperature is correctly identified despite the fact that it is not the most populated structure type in the high-temperature reservoir (Figure 2).

Figure 3 shows a striking agreement between the melting profiles obtained using standard REMD and reservoir REMD simulations. As we noted above, however, analysis of only native populations gives an incomplete view of the composition of an ensemble of structures. To be able to more fully evaluate the ensembles provided by R-REMD, one must compare populations not only of the native conformation but of all accessible states. We selected the ensemble at 350 K for this analysis; the proximity to the T_m makes this an excellent temperature to characterize the ensemble under conditions where native and non-native conformations are well populated. Cluster analysis on the combined set of structures sampled at 350 K in all REMD and R-REMD simulations resulted in 63 clusters with the native conformation being the highest-populated cluster in each simulation (Figure 4). We note that the most populated cluster is different at these temperatures (native at 350 K and non-native at 400 K).

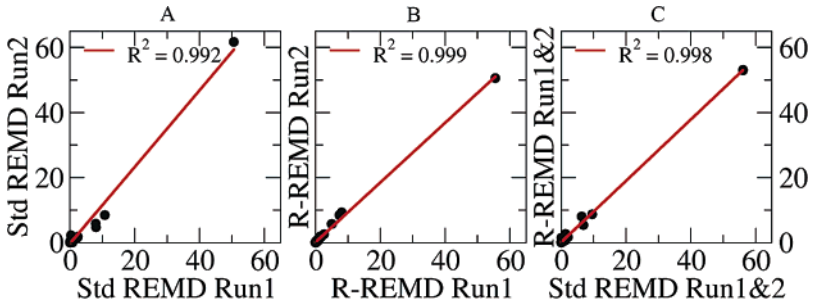

Figure 4. Comparison of the populations of a set of trpzip2 structure types sampled in different simulations. Structure families are defined using the combined set of structures, permitting direct comparison of populations between trajectories. (A) comparison of standard REMD from native vs standard REMD from unfolded. (B) Comparison of R-REMD from native vs R-REMD from unfolded. (C) Comparison of the combined data from standard REMD and the combined data from R-REMD. High correlations were observed in each case ($R^2 \approx 0.99$), and the most populated cluster is the same in all runs. Regression analysis after discarding the most populated cluster results in a similar level of agreement.

Standard REMD simulations starting from different initial conformations show high correlation between cluster populations ($R^2 > 0.99$), suggesting that the ensembles are well converged and the data are suitable as a reference to evaluate R-REMD results (Figure 4A). Similarly both R-REMD simulations starting from different conformations are in excellent agreement with $R^2 > 0.99$ (Figure 4B). Having thus validated the precision of the results from each method, we compare the populations of different structures in the ensemble obtained from standard REMD to that from R-REMD (Figure 4C). The agreement between the two data sets is impressive, with $R^2 = 0.998$ and a slope of 0.932.

The regression data obtained using all clusters may be biased by a single cluster with large population (native). We repeated the regression analysis for the data shown in Figure 4 after removal of this data point, thus comparing the preference to sample the various weakly populated structures in the unfolded state. For all cases, the resulting fit is similar to the original, with correlation coefficients of 0.974 (A), 0.997 (B), and 0.966 (C) between the unfolded ensembles sampled in the REMD and R-REMD simulations. Thus, we conclude that the ensemble obtained from R-REMD is essentially indistinguishable from that obtained using standard REMD, including the relative populations of the various conformations that make up the unfolded state.

Testing the Efficiency of R-REMD. We have demonstrated that R-REMD can produce the same ensembles of structures as standard REMD, validating the accuracy of the approach. We next investigate whether R-REMD offers any advantage over standard REMD in terms of computational cost. To analyze the rate of convergence, the population of native conformation with respect to simulation time was calculated for each simulation and temperature.

Figure 5A shows the native populations vs time for several temperatures in the two independent standard REMD simulations. As expected the values undergo very large fluctuations at the beginning of the REMD run and slowly approach their equilibrium values (obtained by combining the two data sets and discarding a significant amount of data to remove bias from initial conditions as described for Figure 3). After 155 000 exchange attempts (155 ns per replica), populations near the melting temperatures still fluctuate and do not show a flat profile with increasing simulation time. It is interesting

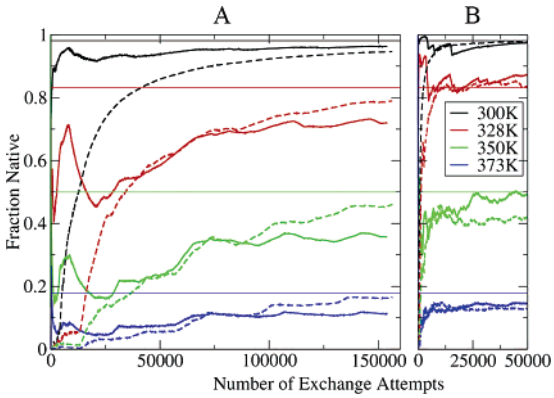

Figure 5. Convergence of native population in standard REMD runs (left) and R-REMD runs (right) vs number of exchange attempts. Solid lines represent simulations starting from native conformation and dashed lines represent simulations starting from unfolded conformations. Thin lines on both graphs represent the average equilibrium values obtained from the standard melting curves (Figure 3). For both graphs, the *x* axis is on the same scale. For standard REMD (left), the results fluctuate at the beginning of the simulations and slowly converge to their equilibrium values. Even though the simulations were extended to 155 000 exchange attempts, the average native populations show about 10% deviation between the two runs at multiple temperatures and plateau values have not been reached. R-REMD simulations (right) converge much faster (~5000−10 000 exchange attempts).

to note that the simulation initiated with all replicas in the native conformation still underestimates the equilibrium native population. Data near the thermal melting transition (where native and non-native conformations are both sampled) is critically important for characterizing the folding landscape. Even at 100 ns per replica, the population values differ significantly from the final values. Importantly, the populations from the two independent simulations provide similar values (i.e., good precision) at times where the population value is dramatically different from the final value (poor accuracy), indicating that precise results for the native population are not a reliable indicator of the overall convergence of the data. As an example, if we perform cluster analysis on the ensembles sampled up to 50 000 exchange attempts, the native population in both simulations

is similar (55% and 60%). However, the correlation coefficient for the populations of unfolded conformations is only 0.796, showing that, even though the largest cluster populations agree with each other, the overall sampling is not complete and the unfolded state and folding landscape may be poorly converged.

In marked contrast to the slow convergence obtained with standard REMD, both R-REMD simulations reach their equilibrium values after only 10 000 exchange attempts and fluctuate around this value for each temperature (Figure 5B). As observed in the melting curves, good agreement between the two methods over the temperature range is observed. The results seem to differ about 7% at 350 K, which is reasonable since the melting transition is sharp around this temperature, and this small difference corresponds to only ~0.16 kcal/mol difference in free energy (49.0 vs 42.3%). Overall, the R-REMD simulations converge to their equilibrium values much faster than standard REMD simulations. Standard REMD simulations have not reached their equilibrium values even at 150 000 exchange attempts (150 ns per replica). In contrast, R-REMD simulations reach their equilibrium values in ~5000−10 000 exchange attempts and remain near these values throughout the remainder of the simulations. This represents an improvement of over an order of magnitude in efficiency with R-REMD as compared to standard REMD.

Up to this point, the R-REMD simulations were compared to standard REMD simulations that employed a much larger temperature range (up to 570 K). For examination of computational efficiency, however, a more direct comparison between standard and R-REMD would involve using the same number of replicas and temperatures for each method. To test this, a new REMD simulation was prepared starting from the same unfolded conformation used for R-REMD but with 5 replicas: 4 matching the temperatures used in the R-REMD run and an additional replica at 400 K. The only difference between this REMD run and the previous R-REMD run is that the 400 K trajectory is a continuous simulation with exchanges that are synchronized with the other replicas instead of being chosen randomly from the pre-generated 400 K structure reservoir used for R-REMD. The native populations versus time for this REMD simulation are shown in Figure S3. These standard REMD simulations with a highest temperature of 400 K converge much more slowly than the original standard REMD which used more replicas covering a wider temperature range. After 180 000 exchange attempts (180 ns per replica), the replicas still did not reach the equilibrium values determined from standard REMD runs, and they also show relatively little progress toward these values. This slow convergence is somewhat unexpected since this REMD simulation was run longer than the cumulative simulation time of our standard MD simulations at 400 K (180 ns per replica vs 152 ns of standard MD), and these standard simulations were shown to be reasonably well converged (Figure 2 and S1). We believe that this difference in convergence between high-temperature MD and REMD demonstrates the effect of "scavenging" of low-energy structures sampled at the highest temperature by the lower temperatures, slowing the convergence of the high-temperature REMD ensemble. This interpretation is consistent with the observation that the lowest temperature converges within ~50 ns to nearly fully native ensemble; once this structure is located at higher T and exchanged to the lowest T, it becomes trapped and no further exchanges take place (and will not until other low-energy basins are located at higher temperatures). Thus the rapid convergence of this low temperature is not an adequate measure for simulation convergence at higher temperatures. As discussed above, temperatures such as 350 K where the native state does not fully dominate the ensemble are likely to be much more useful in characterizing the folding landscape and composition of the unfolded state. The poor convergence of standard REMD at these temperatures and the rapid convergence of R-REMD under otherwise identical conditions confirm that using an equilibrated structure reservoir instead of a synchronous high-temperature replica significantly increases the rate of convergence of REMD simulations.

One remaining question with the R-REMD simulations is how much the convergence rate and final results depend on the composition of the reservoir set. We tested this dependence by repeating the R-REMD run from an initial unfolded ensemble, using only the first half of the original structure reservoir (corresponding to two of the four MD trajectories at 400 K). The resulting pool of 5000 structures had a native population of ~1.5%.The resulting R-REMD melting curve is shown in Figure S4, along with those obtained from standard REMD and R-REMD with the larger reservoir. The thermal stability of trpzip2 in the R-REMD run with the smaller pool is somewhat lower, with ~15 K reduction in the midpoint of the melting transition. This likely results from a lower population of native conformations in the smaller reservoir (~1.5 vs 3%). Even with this much smaller native population in the reservoir, the R-REMD run shows good agreement at the lowest temperatures away from the reservoir, and the native population at higher temperatures is reduced accordingly. The simulations still converge as fast as the R-REMD simulations using the full structure pool (data not shown), suggesting that repeating R-REMD simulations with independent reservoirs would be an excellent approach to validating data convergence.

Testing R-REMD Performance with an Antiparallel β-sheet. To test the efficiency of the R-REMD method on a different and more challenging system, we simulated the peptide dPdP, which has been shown to adopt a 3-stranded antiparallel β-sheet.[49,51] We previously reported results from independent standard REMD simulations starting from fully extended and compact initial conformations.[17] Here, we compare those results to data from new simulations performed using R-REMD, starting from a fully extended conformation.

We employed a single long MD simulation of dPdP to generate the structure reservoir (260 ns, with 5 folding transitions observed). The reservoir was again generated at 400 K, and the native content in the resulting ensemble was 7.7%, in reasonable agreement with data at 399 K in our standard REMD simulations (4.5 and 4.7% in the independent REMD runs). Once again 10 000 structures were selected at equal intervals for use as the structure reservoir. Since dPdP is a larger system than trpzip2, six replicas were

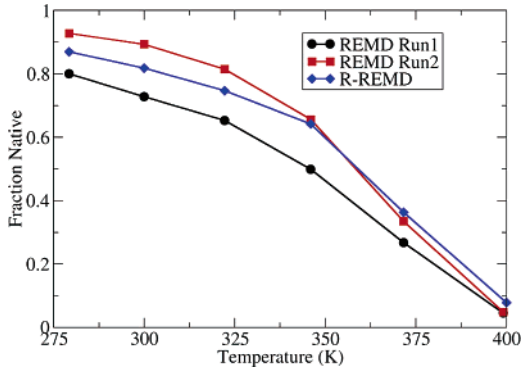

Figure 6. Comparison of dPdP melting curves from standard REMD simulations (black and red) and R-REMD simulation (blue). For standard REMD simulations, data from the first 20 000 exchange attempts were discarded to remove bias introduced by initial conformations. For the R-REMD simulation, the 400 K population reflects the reservoir ensemble.

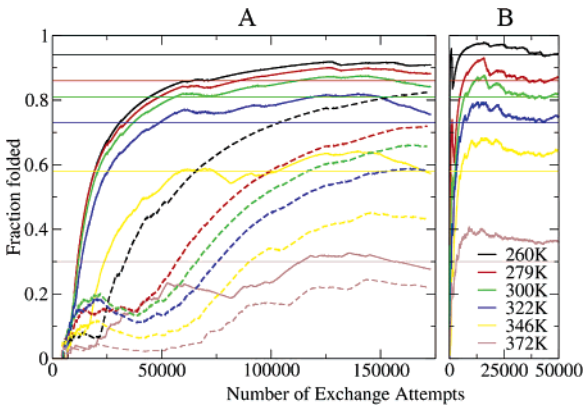

Figure 7. Native fraction vs number of exchange attempts for standard REMD simulation (A) and R-REMD simulation (B). Solid lines in panel A represent simulations starting from compact non-native structure, and dashed lines represent simulations starting from extended conformation. Even after 170 000 exchange attempts, plateau values have not been reached. During R-REMD simulations (B), all replicas converge to their equilibrium values after ~10 000 exchange attempts and show a flat profile thereafter.

used with the same temperature distribution that we employed in the standard REMD simulations resulting in 15−20% exchange ratios between replicas and 14% between highest replica and the 400 K reservoir. No data were discarded, since within 28 exchange attempts every one of the initial fully extended conformations had been exchanged with the reservoir, as expected since the fully extended conformation is energetically less favorable than the MD-generated conformations in the reservoir.

We compare the dPdP melting curves from our standard REMD simulations with the R-REMD results in Figure 6. As we observed with trpzip2, good agreement is obtained between REMD and R-REMD, with the R-REMD melting profile falling within the precision bounds obtained from the two independent standard REMD runs. As we reported previously,[17] these values are in good agreement with experimental observations.

Having confirmed that R-REMD is once again able to accurately reproduce the thermal melting profiles obtained using standard REMD, we evaluated how long it took each simulation to reach these equilibrium values (Figure 7). Even after 170 000 exchange attempts (~2 μs per run) for standard REMD simulations it was not possible to conclude that the simulations were well converged since the populations at some temperatures varied more than 10% and, in many cases, a plateau had not yet been reached.

In contrast with the standard REMD results, dPdP R-REMD simulations reach their equilibrium values within ~10 000 exchange attempts and show an essentially flat profile after that point. Simulations were continued up to 50 000 exchange attempts, with no significant changes for any of the simulated temperatures. The total simulation time used to obtain fully converged ensembles using R-REMD, including reservoir generation (260 ns), was ~320 ns, although we did not test whether a shorter reservoir generation simulation would have been sufficient. On the basis of this time, we estimate that R-REMD is at least 6.4 times more efficient than standard REMD.

Conclusions

We introduced a new variant of the replica exchange method where slow convergence and the high computational cost of REMD have been greatly improved by coupling of the REMD replicas to an ensemble of conformations that is generated in advance, similar in spirit to J-walking schemes. This approach builds on the hypothesis that the main contribution to sampling efficiency during REMD is obtained from the replicas exploring the free-energy landscape at high temperatures. Rather than simulating all replicas during this search process, R-REMD performs the search for alternate local minima *in advance* and subsequently uses a relatively short REMD run to generate accurate Boltzmann-weighted ensembles at other temperatures. An important advantage is that exchanges with the reservoir need not be time-correlated with the replica simulations, permitting REMD replicas to obtain many low-energy (such as native) conformations from a smaller number of folding events; this is not possible with standard REMD, which may a contributing factor in slow convergence.

In our current implementation in Amber, we employed a relatively small ensemble of 10 000 structures in the reservoir. Although successful exchanges would, in principle, involve placing the structure from the MD replica into the reservoir and removing the previous reservoir structure, we made the approximation that the reservoir is large enough that these changes would not affect the reservoir ensemble, and therefore, reservoir was not modified for computational convenience. As a result, the present implementation in Amber does not strictly obey detailed balance conditions; however, the resulting ensembles are in very good agreement with the standard REMD data. Furthermore, the use of a larger reservoir or true exchange of conformations could be readily accommodated in future implementations, and we note that the method itself is rigorous.

We tested R-REMD by comparing to standard REMD results for two systems, a β-hairpin and a three-stranded

β-sheet, under conditions in which the standard REMD data were in good agreement with experimental observations. We find that the thermal melting profiles obtained from R-REMD simulations were highly accurate compared to those of standard REMD, as expected because of the lack of any approximations in development of the method. Furthermore, excellent agreement was noted between the compositions of the structure ensembles obtained from standard REMD and R-REMD, including very high correlations between the two methods for the populations of native and non-native conformational families.

To summarize the efficiency comparisons, standard REMD simulations with 14 trpzip2 replicas were run for 155 ns per replica from two initial conformations resulting in a cumulative simulation time of ~4.3 μs, and they still did not fully converge. The R-REMD simulations were run using 4 replicas and two initial conformations, and both runs reached their equilibrium values in under 10 ns per replica (40 ns total). Generation of the reservoir does require additional computational effort that must be included in the comparison. In the present case, four simulations of ~40 ns were employed (152 ns, almost as long as each replica in the REMD simulations but importantly only 1 temperature is needed for R-REMD). The ability to use multiple simulations provides the reservoir generation with parallel efficiency comparable to the REMD simulations. Thus the cumulative simulation time for R-REMD *including the reservoir generation* is about 232 ns, approximately 19 times more efficient than the less well converged 4.3 μs standard REMD simulation. These values result from comparison to a primitive standard REMD implementation in Amber. Future studies will compare the efficiency of R-REMD to other, more recent REMD variants, as well as investigate how the R-REMD approach can be further optimized

Acknowledgment. C.S. thanks Adrian Roitberg for helpful discussions. C.S. gratefully acknowledges financial support from the NIH (GM6167803) and DOE (DE-AC02-98CH10886) and significant supercomputer time at NCSA (NPACI MCA02N028).

Supporting Information Available: Figures S1−S4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tai, K. Conformational sampling for the impatient. *Biophys. Chem.* **2004**, *107* (3), 213−220.

(2) Roitberg, A.; Simmerling, C. Special issue: Conformational sampling. *J. Mol. Graphics Modell.* **2004**, *22* (5), 317−317.

(3) Smith, L. J.; Daura, X.; van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simluations. *Proteins: Struct., Funct., Genet.* **2002**, *48* (3), 487−496.

(4) Geyer, C. J.; Thompson, E. A. Annealing Markov-chain Monte-Carlo with applications to ancestral inference. *J. Am. Stat. Assoc.* **1995**, *90* (431), 909−920.

(5) Hukushima, K.; Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65* (6), 1604−1608.

(6) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281* (1−3), 140−150.

(7) Swendsen, R. H.; Wang, J. S. Replica Monte-Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57* (21), 2607−2609.

(8) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. Monte Carlo study of the interacting self-avoiding walk model in three dimensions. *J. Stat. Phys.* **1996**, *82* (1−2), 155−181.

(9) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.

(10) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21* (6), 1087−1092.

(11) Feig, M.; Karanicolas, J.; Brooks, C. L. MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22* (5), 377−395.

(12) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins: Struct., Funct., Genet.* **2001**, *42* (3), 345−354.

(13) Garcia, A. E.; Sanbonmatsu, K. Y. α-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (5), 2782−2787.

(14) Karanicolas, J.; Brooks, C. L. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (7), 3954−3959.

(15) Kinnear, B. S.; Jarrold, M. F.; Hansmann, U. H. E. All-atom generalized-ensemble simulations of small proteins. *J. Mol. Graphics Modell.* **2004**, *22* (5), 397−403.

(16) Pitera, J. W.; Swope, W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (13), 7587−7592.

(17) Roe, D. R.; Hornak, V.; Simmerling, C. Folding cooperativity in a three-stranded β-sheet model. *J. Mol. Biol.* **2005**, *352* (2), 370−381.

(18) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042−6051.

(19) Zhou, R. H.; Berne, B. J.; Germain, R. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (26), 14931−14936.

(20) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109* (16), 8220−8230.

(21) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.

(22) Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **2002**, *117* (15), 6911−6914.

(23) Rathore, N.; Chopra, M.; de Pablo, J. J. Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* **2005**, *122* (2), 024111.

(24) Jang, S. M.; Shin, S.; Pak, Y. Replica-exchange method using the generalized effective potential. *Phys. Rev. Lett.* **2003**, *91* (5), 058305.

(25) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J. Chem. Phys.* **2003**, *118* (14), 6664−6675.

(26) Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **2000**, *329* (3−4), 261−270.

(27) 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.

(28) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (24), 13934−13939.

(29) Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L. The unfolded state of the villin headpiece helical subdomain: Computational studies of the role of locally stabilized structure. *J. Mol. Biol.* **2006**, *360* (5), 1094−1107.

(30) Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct., Funct., Genet.* **2002**, *46* (2), 225−234.

(31) Zhang, W.; Wu, C.; Duan, Y. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* **2005**, *123*, 154105.

(32) Zuckerman, D. M.; Lyman, E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* **2006**, *2*, 1200−1202.

(33) Zhang, J.; Qin, M.; Wang, W. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. *Proteins: Struct. Funct. Bioinformatics* **2006**, *62* (3), 672−685.

(34) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. α-Helix peptide folding and unfolding activation barriers: A nanosecond UV resonance Raman study. *J. Am. Chem. Soc.* **1999**, *121* (35), 8074−8086.

(35) Matagne, A.; Jamin, M.; Chung, E. W.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. *J. Mol. Biol.* **2000**, *297* (1), 193−210.

(36) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **1997**, *390* (6656), 196−199.

(37) Oliveberg, M.; Tan, Y. J.; Fersht, A. R. Negative Activation Enthalpies in the Kinetics of Protein-Folding. *Proc. Natl. Acad. Sci.* U.S.A. **1995**, *92* (19), 8926−8929.

(38) Segawa, S. I.; Sugihara, M. Characterization of the transition-state of lysozyme unfolding. 1. Effect of protein solvent interactions on the transition-state. *Biopolymers* **1984**, *23* (11), 2473−2488.

(39) Cavalli, A.; Ferrara, P.; Caflisch, A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins: Struct., Funct., Genet.* **2002**, *47* (3), 305−314.

(40) Ferrara, P.; Apostolakis, J.; Caflisch, A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B* **2000**, *104* (20), 5000−5010.

(41) Frantz, D. D.; Freeman, D. L.; Doll, J. D. Reducing quasi-ergodic behavior in Monte-Carlo simulations by J-walking−Applications to atomic clusters. *J. Chem. Phys.* **1990**, *93* (4), 2769−2784.

(42) Zhou, R. H.; Berne, B. J. Smart walking: A new method for Boltzmann sampling of protein conformations. *J. Chem. Phys.* **1997**, *107* (21), 9185−9196.

(43) Andricioaei, I.; Straub, J. E.; Voter, A. F. Smart darting Monte Carlo. *J. Chem. Phys.* **2001**, *114* (16), 6994−7000.

(44) Brown, S.; Head-Gordon, T. Cool walking: A new Markov chain Monte Carlo sampling method. *J. Comput. Chem.* **2003**, *24* (1), 68−76.

(45) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Resolution exchange simulation. *Phys. Rev. Lett.* **2006**, *96* (2).

(46) Lyman, E.; Zuckerman, D. M. Resolution exchange simulation with incremental coarsening. *J. Chem. Theory Comput.* **2006**, *2* (3), 656−666.

(47) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. A.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(48) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci.* U.S.A. **2001**, *98* (10), 5578−5583.

(49) Schenck, H. L.; Gellman, S. H. Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. *J. Am. Chem. Soc.* **1998**, *120* (19), 4869−4870.

(50) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127−6129.

(51) Syud, F. A.; Espinosa, J. F.; Gellman, S. H. NMR-based quantification of β-sheet populations in aqueous solution through use of reference peptides for the folded and unfolded states. *J. Am. Chem. Soc.* **1999**, *121* (49), 11577−11578.

(52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of Cartesian equations of motion of a system with constraints−Molecular-dynamics of *N*-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.

(53) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.

(54) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* **2003**, *24* (1), 21−31.

(55) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246* (1−2), 122−129.

(56) Ponder, J. W.; Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8* (7), 1016−1024.

(57) Simmerling, C.; Elber, R.; Zhang, J. MOIL-View, a program for visualization of structure and dynamics of biomolecules, and STO, a program for computing stochastic paths. In *Modelling of Biomolecular Structures and Mechanisms*, Proceedings of the 27th Jerusalem Symposium on Quantum Chemistry and Biochemistry, Jerusalem, Israel, May 23−26, 1994; Pullman A., Jortner, J., Pullams, B., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp 241−265.

CT600263E