JCTC Journal of Chemical Theory and Computation

Convergence and Heterogeneity in Peptide Folding with Replica Exchange Molecular Dynamics

Edmund Lin and M. Scott Shell*

Department of Chemical Engineering, University of California Santa Barbara, 552 University Road, Santa Barbara, California 93106-5080

Received March 13, 2009

Abstract: Replica exchange molecular dynamics (REMD) techniques have emerged as standard tools for simulating peptides and small proteins, in part, to evaluate the accuracy of modern classical force fields for polypeptides. However, it often remains a challenge to unambiguously discriminate force field flaws from simulations that do not reach equilibrium convergence. Here, we examine closely the convergence behavior of REMD runs for 14 test alpha and beta peptide systems, using an AMBER force field with a generalized Born/surface area implicit solvation model. Somewhat surprisingly, we find that convergence times can be quite large compared to the time scales reached in many earlier REMD efforts, with some short peptides requiring up to 60 ns of run time (per replica). Moreover, we detect a high degree of run-to-run heterogeneity, finding that REMD runs of the same peptide seeded with different initial velocities can exhibit a range of fast- and slow-folding behavior. By increasing the number of swap attempts per REMD cycle, we are able to reduce heterogeneity by diminishing the presence of slower-folding trajectories. Finally, we notice that convergence often can be signaled by a spike in the population of the most populated configurational cluster — a metric that is independent of the native structure. These results suggest that the systematic application of long runs, multiple trials, and convergence indicators may be important in future folding studies and in force field development efforts.

Introduction

The prediction of the three-dimensional native structures of proteins from their sequences remains a prominent challenge. It has been particularly challenging to make such predictions without bioinformatics scoring functions or templates, and instead using purely "physics-based" methods; these techniques involve canonical sampling and classical physiochemical energy functions and solvation models designed to reproduce true interatomic energetics. Such approaches have met two limitations: (1) computational challenges in sampling the vast conformational spaces and concomitant long folding time scales of proteins, and (2) potential deficiencies in current force field parametrizations. In order to delineate among these potential sources of error, it is critical to identify whether a simulation has reached equilibrium convergence, as even accurate force fields will not correctly fold a protein if the simulation time scale is less than the intrinsic folding time. Therefore, the ability to assess and reach convergence in folding simulations is of major importance.

Two complementary approaches have emerged as standards for increasing simulation convergence and efficiency. First, implicit solvation models reduce the computational overhead of explicit water simulations and remove solvent viscosity that slows exploration of the conformational landscape. In particular, ongoing efforts have refined fast generalized Born/surface area (GBSA) techniques that generate approximate solutions to the continuum solvent electrostatics[1] with a nonpolar term proportional to solute surface area.[2] Second, the parallel replica exchange molecular dynamics (REMD) technique of Sugita and Okamoto speeds convergence relative to brute force molecular dynamics.[3] In REMD, a number of copies ("replicas") of a system are evolved simultaneously, each at a different temperature

* Corresponding author. E-mail: shell@engineering.ucsb.edu.

10.1021/ct900119n CCC: $40.75
Published on Web 07/24/2009

spanning that of interest (e.g., physiological temperature) to a heated state more conducive to traversing free energy barriers. Periodic swaps of neighboring replicas enable conformations to heat up and cool down and are performed in such a way as to maintain an overall Boltzmann-weighted ensemble at each temperature.

In this work, we assess the convergence properties of REMD for a number of peptides modeled with an atomic force field and GBSA solvation. We demonstrate that, even for very short systems, convergence times can be notably longer than those suggested by earlier studies. Moreover, we detect substantial run-to-run heterogeneity in simulations of the same helical peptide, with trajectory folding rates varying by up to a factor of 4−5. We then perform folding simulations of a variety of α-helical and β-hairpin peptides of 11−25 amino acids in length, finding convergence times as great as 60 ns (4.0 aggregate CPU-years). Our results suggest several strategies for assessing and accelerating convergence, and they emphasize the importance of well-converged simulations in the evaluation of force field performance.

Our work follows a long history of computational folding studies. Early efforts used Monte Carlo methods[4−6] and single trajectory[7−12] and multitrajectory molecular dynamics with massive distributed computing[13−15] to examine small peptide systems with well-defined secondary structures and proteinlike folding behavior.[16] Initial applications of the REMD technique employed explicit[17−19] and implicit water[20] simulations that were limited to very short times (≤10 ns). Some of these studies suggested substantial force field sensitivity[8,11,19,21] and raised questions about the balance of α/β energetics, particularly in the backbone torsions.[6,13,17,22]

In recent years, however, REMD simulations using GBSA solvation have typically been performed for 10s or 100s of ns, and many of the earlier studies would now be considered much shorter than modern standards. Rao and Caflisch, for example, performed a 1 μs simulation of a 20-residue β-sheet peptide and observed 312 folding and unfolding events.[23] In a series of separate force-field refinement studies, Jang et al. used 20−130 ns runs to fold a variety of α, β, and α/β peptides up to 28 residues in length.[24−28] In particular, their work demonstrated the need for long runs in converging slow or frustrated folders;[25−27] their refined force field[25] (a modification of AMBER 99[29]) ultimately successfully folded de novo a number of systems to sub 3 Å accuracy.[27] Lei et al. folded the villin headpiece, a 35-residue helical miniprotein, to <2 Å accuracy using 200 ns REMD simulations and the AMBER 03 force field. The same group recently folded the 45-residue helical protein A to high accuracy, requiring a 200 ns effort.[30]

The issues of efficiency and convergence in biomolecular simulations have been addressed in a number of ways. Several groups have argued that REMD speeds convergence relative to conventional molecular dynamics simulations for the equivalent amount of computational effort.[23,30−33] Comparing conventional to replica exchange simulations, authors have found speedup factors of 14−72,[31] 4−8,[32] and ∼5,[30] generally decreasing with the size of the protein. Daura and co-workers suggested that convergence could be signaled by a leveling in the number of clusters taken from a sliding-window configurational clustering analysis,[34,35] and Lyman and Zuckerman later extended this approach with a faster clustering algorithm and the additional consideration of cluster populations.[36] Recently, Abraham and Gready developed a simple model for assessing the "mixing efficiency" of REMD simulations related to the rate at which replicas are able to decorrelate from temperatures during exchanges.[37] This work suggests an absolute minimum simulation length in terms of system size and number of replicas, which is computed under the assumption that successive exchange attempts are completely decorrelated; while their estimated minimum times are actually small relative to typical efforts in the literature, correlations are likely to greatly increase such estimates.

Recent simulation efforts have attempted to clarify long-standing issues about potential flaws in modern force fields, in part, by achieving better simulation convergence than early studies. Studies have addressed the accuracy of backbone torsional energetics related to the appropriate balance between α-helical and β-sheet secondary structure motifs. Several groups have performed direct comparisons of simulations with infrared spectroscopy and NMR experiments.[38−40] Importantly, these efforts identified an increased helical propensity in early force fields like AMBER 94[38,39] and even suggested persistent problems with more recent parametrizations,[39] an issue that remains, perhaps in part, due to a greater focus on helical miniproteins in recent extensive REMD folding efforts. Still, several groups have pursued modification of the backbone torsional terms to improve accuracy and reduce helicity.[24,25,38,41,42] Hu et al. also showed that a quantum mechanical treatment improved the prediction of Ramachandran distributions in capped monopeptides relative to several classical force fields and suggested problems with the latter's parametrization by in vacuo rather than solvated ab initio calculations.[43] Some more recent parametrizations have attempted to address this problem with ab initio calculations designed to better mimic solvent polarization.[44]

GBSA implicit solvation models have also been scrutinized as a potential source of error. Earlier studies by Zhou and co-workers found that free energy surfaces of a hairpin peptide were distorted by implicit solvation when compared to explicit.[45,46] Several groups have suggested that a major contributor to this problem has been salt bridges, which are overstabilized by the GBSA model.[46−49] In addition, Roe et al. have found that implicit solvation increases the helical propensity of force fields through its effect on backbone hydrogen-bonding strengths.[49] Two approaches have been pursued to correct these flaws. Okur et al. propose hybrid explicit/explicit REMD simulations that use a solvation shell of explicit water near a protein with an implicit approach beyond. Groups have also pursued modification of the intrinsic GB radii of various atom types.[25,48] In both approaches, the authors found improved agreement with experimental structures and interaction strengths.

Still, it remains challenging to unambiguously decouple force field assessment and simulation convergence in REMD folding simulations. In particular, for force field studies that

do not yield good agreement with experiments, it is essential to rule out the possibility of unconverged results and too short run times. Three issues potentially affect this consideration: (1) **heterogeneity** − Molecular dynamics studies have shown highly heterogeneous folding behavior in even fast-folding peptides.[50] These results suggest REMD simulations can also be characterized by significant run-to-run variances. What is the magnitude of this heterogeneity, to what extent are multiple trials required to attain an accurate picture, and can such variances be reduced? (2) **fluctuations** − Short peptides typical of the lengths currently amenable to direct folding studies are likely to experience significant conformational fluctuations when compared to globular proteins.[36] Studies already suggest that converged simulations typically give average root-mean-square deviation (rmsd) values in the range of 2−4 Å.[12,17,27,30,33,36] What is a reasonable baseline for comparing structures in these simulations when fluctuations are present? (3) **convergence times** − Some studies have found REMD folding times that significantly exceed the short, sub 10 ns run durations of early studies.[25−27] What is the range of convergence times that one might expect for short peptides, and are there any metrics, independent of knowledge of the native structure, that can signal convergence?

It is the goal of this work to provide some perspectives on these questions using extensive REMD simulations. Here, we assess the behavior of REMD simulations performed with a modern and classical interatomic force field in folding a wide range of short peptides taken from the PDB. In particular, we focus on the magnitude, breadth, and heterogeneity of convergence time scales across the spectrum of sequences studied, performing long simulation runs where necessary to provide systems with ample sampling time. For a force field, we use an AMBER variant with an implicit solvation model that we and our co-workers previously found to correctly stabilize the folded structures of small peptides and proteins.[51−53] We describe two studies, a detailed, multitrial examination of the short helical EK peptide and a broader investigation of convergence in a variety of α and β peptide systems.

Methods

We examined 11 peptides taken from the Protein Databank (PDB) and 3 suggested by earlier folding studies.[19,21,26,27] For those taken from the PDB, we filtered for structures that were classified as peptide in the SCOP (structural classification of proteins) scheme,[54] had well-defined secondary structure elements, were solved under normal aqueous solution conditions, and had, at most, a single disulfide bond. Among the other three peptides, the helical EK peptide has been studied by Baldwin and co-workers[55] using circular dichroism and was found to have partial helicity (∼40%) at 0 °C; we took its native structure to be an energy-minimized ideal helix for this sequence. The 15-β peptide was engineered by ref 56 and studied earlier by Kim et al.;[27] we obtained the unpublished structure from the authors in ref 56. The C-peptide, a helix from the C-terminal region of ribonuclease A originates from the work of Baldwin and co-workers,[57] and it has been extensively studied by simulation;[4,19,21,26,47] we take its native structure from the crystal structure of ribonuclease A.

We simulated each peptide using REMD. Our replica temperatures spanned 270−600 K and swaps were attempted every 20 ps of MD simulation. The number of replicas varied from 15−25 and were adjusted so that swap attempts were accepted on average with ∼40−50% probability. MD trajectories were produced using the sander program in the AMBER 9 software suite, and replica swapping and data collection were performed using a custom Python wrapper. Unless indicated otherwise in the PDB files or reference experimental data, all peptides were simulated without capping groups (i.e., with charged termini). Simulations were initiated with peptides in extended conformations, and trajectory frames were recorded every 1 ps. Any disulfide bonds present in the native structure were applied to the peptides at the start of the simulations.

To reach the target temperature in each replica, we used Anderson thermostats with massive collisions after every swap cycle, similar to the approach in refs 51−53 and 59. The choice of the Anderson thermostat is motivated by the fact that, unlike the Berendsen scheme, it rigorously attains a limiting stationary distribution of states that is canonical, as shown in the original derivation in ref 60. In a REMD setting, each between-swap MD trajectory, starting with Boltzmann-randomized initial velocities, is equivalent to a hybrid Monte Carlo move with 100% acceptance (due to energy conservation);[61] as a result, the Markovian evolution due to both swaps and the Andersen thermostats preserves a limiting canonical distribution in each replica and maintains detailed balance. To verify that the swap/collision frequency does not affect our results, we performed comparison simulations for the EK system with a 1 ps frequency and found no qualitative differences in the results from the 20 ps case. In particular, the convergence times and long-time equilibrium ensembles were the same.

Interatomic interactions were modeled using the AMBER ff96 force field[62] with the implicit solvation model denoted "igb = 5" of Onufriev, Bashford, and Case.[63] In a previous study, we and our co-workers found that this force field stabilized well the folds of a number of small peptides[52] and, when combined with a mechanism-based conformational search algorithm, folded several small proteins.[51,53] Some earlier efforts found in short simulations that the ff96 force field did not adequately stabilize helical conformations,[8,39,64,65] although a few groups found it to exhibit α/β balance[19,21] and perform reasonably well with implicit GBSA models.[46] We emphasize that, although we show later the performance of this force field is reasonable, provided simulations are carried out long enough, our study here is not primarily a force field test. Rather, our main focus is the general features of convergence in REMD simulations. In addition to the ff96 force field, we also applied the fix proposed by Simmerling and co-workers[48] that diminishes the intrinsic hydrogen radii on basic residues, in order to bring implicit solvation ion-pairing stabilities in better line with those found in atomistic simulations.

In our initial studies, we considered two variants of the REMD approach. In the first, each cycle, corresponding to

Table 1. Peptides Studied in This Work

peptide/ PDB code	number of residues	secondary structure	disulfide bonds	number of replicas	REMD run length (ns)
1GM2	11	beta	CYS2-CYS10	15	60
C peptide	13	alpha		20	20
EK peptide	14	alpha		20	100 × 5
1O8Y	14	beta	CYS11-CYS3	20	40
15-beta	15	beta		20	20
1E0Q	17	beta		20	20
2I9M	17	alpha		20	80
2JNI	21	beta	CYS3-CYS20	20	40
2JSB	21	beta	CYS3-CYS20	20	40
1WN8	22	alpha		20	60
1B5N	23	beta	CYS7-CYS19	25	100
2EQH	23	beta	CYS1-CYS19	25	200
2DJ9	23	beta	CYS7-CYS19	25	200
1LFC	25	beta	CYS3-CYS20	25	200

one round of MD in each replica followed by swaps, entailed a number of swap attempts equal to the number of temperature neighbors. That is, each possible swap between temperature neighbors was attempted once. In practice, a list of these possible swaps was generated, randomized in order, and then considered in serial succession until the list was exhausted. In a second approach, we increased the number of swaps per temperature neighbor pair per cycle to five. That is, we repeated the aforementioned procedure five times before a second cycle commenced. Since these additional swaps are performed using the same Boltzmann-like acceptance criterion as in the usual REMD approach, it continues to preserve the correct stationary distribution at each temperature.

Simulations of each peptide were analyzed along 5 ns sliding windows across the trajectories; averages corresponding to the window midpoint were computed as a function of time for several quantities. Namely, we computed the rmsd of each peptide backbone from the native or presumed native structure and the population of the most dominant structure after configurational clustering. Clustering was performed using a modified k-means algorithm and a cluster member cutoff of 2 Å rmsd. Populations were determined from the fraction of total trajectory frames that were found to be a member of each cluster, and the structure representing a cluster was taken from the member closest to the cluster centroid. In some cases, we also analyzed the population of intrabackbone hydrogen bonds, for which we used the DSSP criterion[66] to define the hydrogen-bonding conditions.

Results and Discussion

Folding Heterogeneity in the EK Peptide. Our initial efforts sought to characterize in detail the REMD folding behavior of one peptide, the 14-residue helical EK peptide.[55] For this peptide, we performed five independent 100 ns runs, each starting from an extended conformation but with different initial Boltzmann-distributed atomic velocities. All of the analyses discussed below pertain to the lowest-temperature replica at 270 K. Times discussed are on a per replica basis, not aggregate over all replicas.

The most conspicuous, and perhaps surprising, observations in these results is that, even for this small a peptide, the folding runs reveal a very high degree of heterogeneity. Figure 1 shows the rmsd (with respect to an ideal helix of

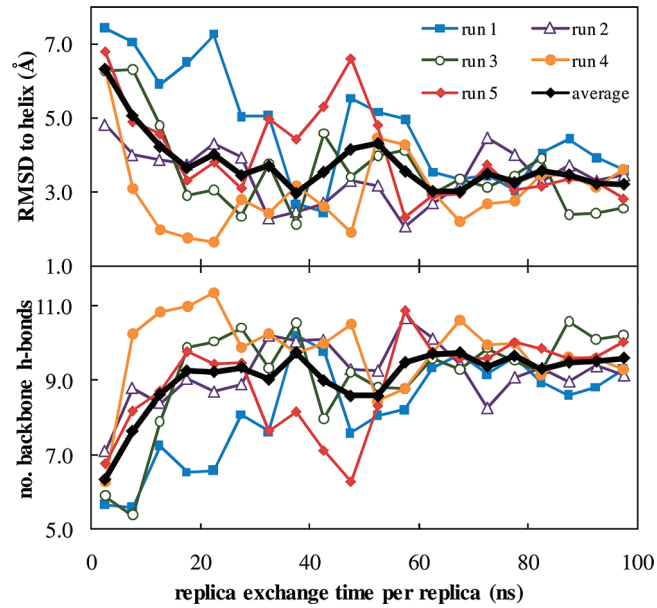

Figure 1. REMD folding of the EK peptide from an extended conformation. (Top) rmsd to an ideal helix as a function of simulation time for five independent runs and their average. (Bottom) Corresponding average number of backbone hydrogen bonds. Each point represents an average over a window of 5 ns centered around it.

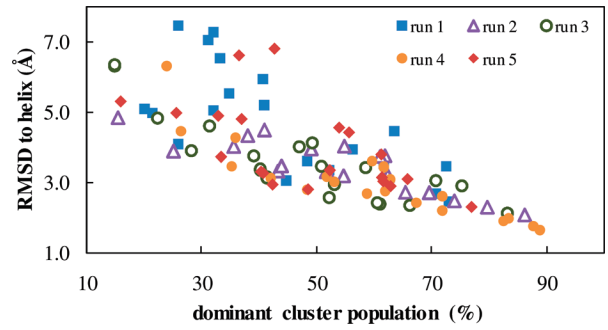

Figure 2. Relationship between rmsd and dominant cluster population for the EK peptide. Each point corresponds to a 5 ns sliding window during the simulation trajectory: the *x*-axis contains the population of the largest conformational cluster extracted from snapshots during that window, and the *y*-axis contains the average rmsd for the same snapshots.

identical length) and the hydrogen-bond population as a function of time in the simulation. Run one remains largely unfolded (rmsd > 4.0 Å) for at least the first 30 ns of the simulation, whereas run four rapidly approaches the native state and achieves an rmsd of 2.0 Å in roughly 10 ns. Run five, on the other hand, initially appears to fold at an average rate but experiences a large increase in rmsd at 30 ns that persists for ∼20 ns before the system returns to its final average value. An analysis of the run five trajectory reveals that the system folds to a metastable structure during the intermediate times, a β-hairpin structure (shown in Figure 2). In this run, the hairpin forms in higher-temperature replicas and then swaps to the lowest before the system can locate the more favorable helical structure. Overall, despite the small size of this peptide, the collection of runs manifests fast-folding, slow-folding, and misfolding trajectories.

Though the runs differ widely in their individual folding, their mean behavior exhibits some notable trends. Averaged over all five runs, the rmsd versus time curve is well described by a first-order process and exponential decay with a time constant of ~8 ns. There is a very strong correlation between rmsd and hydrogen bonding, with low rmsd values almost always paralleling the number of backbone hydrogen bonds formed, in both the average and five individual runs. This result is quite expected for a peptide of this size with a helical structure and geometry which maximizes the amount of intrapeptide hydrogen bonds.

Moreover, we find that convergence in each of the runs can be signaled by a conformational clustering analysis, an indicator that is independent of the native structure. We divided each run into 5 ns windows and performed a clustering of the corresponding trajectory frames; the population of the top cluster was taken as an indicator of the degree of conformational stability within each window. As shown in Figure 2, these dominant cluster populations exhibit a strong correlation with the respective average rmsd of the windows. That is, parts of the simulation that experience less fluctuations appear to be closer to the native structure. Remarkably, the results from the five runs all fall on the same line at low rmsd values and high cluster populations, despite heterogeneities among them. These results suggest the exciting prospect that convergence in each run can be assessed without knowing the native structure but by examining conformational populations. In all runs, a top cluster population of greater than 70% indicates a structure that is of typical rmsd in the equilibrated portions of the simulations. This finding supports the earlier arguments of Daura and co-workers[34,35] and Lyman and Zuckerman,[36] that clustering provides a convenient convergence signal.

How do the peptides reach a folded state in each run? To quantify this behavior, we define a folding event as a point in time at which a folded structure first appears in any of the replicas, given that no folded structure exists in any replica for the previous 100 ps of REMD time. Here, we define "folded" as a snapshot that has an rmsd of 2 Å or less from the ideal helical structure. The distribution of such folding events over the replica temperatures and for the different runs is shown in Figure 3. For each of the five runs, folding events happen in replicas spanning the lowest temperature (270 K) to slightly above the folding temperature (T_f ~ 350 K, estimated from the temperature-dependence of the radius of gyration). Moreover, we find that multiple folding events occur in each run due to the fact that folded structures do not always persist through further time evolution. It is interesting to note that the fast folding trajectory (run four) has a higher frequency of folding events near and just below the folding temperature, relative to low temperatures. More strikingly, the misfolding trajectory (run four) has a pronounced peak in folding events near T_f. Both results might suggest that the conformational transitions in these systems stem from higher-temperature events that then propagate to the lowest-temperature replica via swaps. The remaining runs all tend to have higher probabilities of folding events at the lower-temperature replicas. Overall, however,

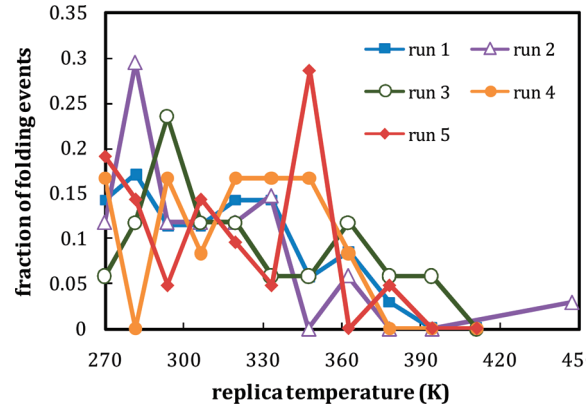

Figure 3. Distribution of folding events in replica space for each EK peptide run. A folding event is defined as a point in which a ≤2 Å rmsd structure is first reached in any replica given that none exists for the previous 100 ps of REMD time.

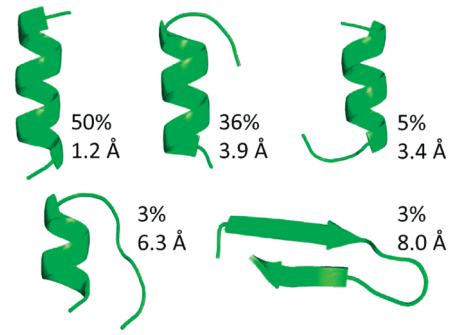

Figure 4. Dominant conformational clusters for the EK peptide. Structures are produced by conformational clustering of the final 40 ns of each of the five independent runs, for a total of 200 ns REMD time. Percentages correspond to the fraction of trajectory snapshots (spaced every 1 ps apart) contained within each cluster group. Rmsd numbers are taken with respect to an energy-minimized ideal helix.

the distributions for all of the runs demonstrate heterogeneity in that folding occurs at different temperatures.

Even after each of the runs converges to its limiting behavior within the 100 ns time frame, the data show that significant conformational fluctuations remain. All five runs appear to have reached a steady-state by a time of ~60 ns, beyond where their rmsd values fluctuate around an average of 3.2 Å. Figure 4 shows the structures sampled during these fluctuations and extracted using a clustering analysis of the aggregate final 40 ns from each of the runs (for a total of 200 ns). After equilibration, the peptide spends half of its time near a structure that is very close to the reference ideal helix, at 1.2 Å rmsd. At other times, this system fluctuates to other structures in which either the C-terminal (39% of the time) or N-terminal (5% of the time) parts of the chain partially unfold. In these cases, while the peptide still contains significant helical content, its rmsd from the ideal structure is rather high at 3.4−6.4 Å. The clustering analysis also shows that a very small fraction (1%) of fluctuations locate the β-hairpin structure that was found to be in a misfolded state in the run five trajectory. However, the majority of the conformational fluctuations are centered around a helical

structure, and it is these that contribute to the equilibrium average rmsd of 3.3 Å.

Are these conformational fluctuations representative of the true behavior of the EK peptide found in experiments? Circular dichroism measurements on this peptide reveal only partial helicity near 270 K, which is around a level of 40% when interpreted with a simple Zimm−Bragg model.[55] It is challenging to determine whether these experiments indicate that the peptide fluctuates between fully and partially folded conformational states or folds to a single, partially helical structure. However, experiments do show the lack of a sharp unfolding transition as urea is added to the EK system, particularly relative to longer helical peptides; this peptide's helicity smoothly and continuously decreases with an increasing urea concentration. Such a low degree of folding cooperativity might suggest that the folded state is not a single structure but evolves from an ensemble of conformations. Thus, it seems reasonable that a peptide of this short size would manifest some degree of structural variety in solution. Moreover, our computed average rmsd values are consistent with those typically found in simulations of short peptides.[12,17,27,30,33,36] A direct comparison with experiment remains difficult, owing to challenges in bridging the circular dichroism measurements with simulation observables.

It is worthwhile to comment on the role of the force field and the implicit solvation model in these results. Earlier studies with this force field, AMBER ff96, suggested that it overstabilized β-structures,[8,39,64,65] although a few later investigations by us[52,53] and others[19,21,51] found it to exhibit balance between helix and sheet. One potential factor here may be our use of an implicit solvation model, which may favor α-helical structures relative to explicit water simulations via its stabilizing effect on backbone hydrogen bonding.[49] In that sense, it is possible that a β-propensity in the force field is being corrected fortuitously by an α-propensity in the solvation model.

Alternatively, this work shows that the times scales required for proper folding and convergence greatly exceed those that were attained in earlier studies. Moreover, we find that investigations based on the results of single runs are potentially suspect; in the run five trajectory, for example, a short run would have suggested a strong β-propensity in the force field due to the intermediate formation of a misfolded hairpin structure, when ultimately this state represents only 1% of the equilibrium conformational ensemble. Thus, the possibility remains that early REMD folding studies might have been biased by short simulation times.

Relative to explicit water simulations, the use of an implicit solvation model can contribute to both a speeding up and slowing down of convergence in our runs. On one hand, the lack of any viscosity facilitates a faster exploration of conformational space, and under the premise that an implicit model maintains the same folding free energy, one would expect convergence to be obtained significantly faster with an implicit rather than explicit solvation approach. On the other hand, implicit models have been found to overstabilize ion-pairing interactions, which can lead to the formation of kinetic trap conformations driven by nonnative salt bridges.[46−49] The EK peptide contains four potential salt bridges, three of which

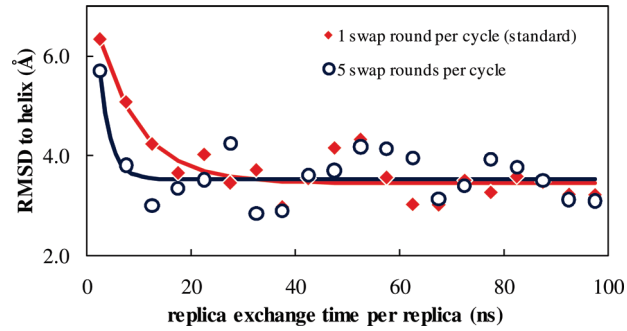

Figure 5. Effect of increasing the number of swaps per round. Each of the two sets of points represents the average over five independent trajectories. The lines correspond to exponential fits of the form rmsd = $C_1 + C_2 \exp(-t/\tau)$. The red diamonds are identical to the average results reported in the top panel of Figure 1.

can be formed in the native structure. Our analysis shows that, beyond the initial 10 ns of the trajectories, the peptide has, on average, 0.4 salt bridges. We do not find a significant correlation between salt bridge population and convergence behavior, and given the relatively small value of this number, we conclude that these interactions are not slowing convergence for this particular case.

We investigated the possibility of accelerating convergence by increasing the number of swap attempts performed per each REMD cycle. Swaps generally require little computational overhead relative to the molecular dynamics trajectories, and thus, this strategy represents a simple and efficient way to improve performance. In principle, the number of swaps per cycle should have no limiting effect on the peptide configurational ensemble, as these preserve the detailed balance equation, but increasing their number might accelerate convergence by permitting higher-temperature replicas that correctly fold to move more quickly to lower ones.

Figure 5 compares the base case (one swap per neighbor pair per cycle) to the average behavior of five runs in which we increase the number of swap attempts 5-fold. The increase in swaps clearly accelerates mean folding; a detailed analysis shows that the individual runs become less heterogeneous, with the swap increase apparently filtering out slow- and misfolding REMD trajectories. In fitting this new case with an exponential, the time constant is nearly a factor of 3 less than that of the original REMD runs at 3 vs 8 ns. Thus, increasing the number of swaps appears to be a promising strategy for improving REMD convergence by removing slow-folding trajectories with minimal additional overhead.

Convergence for Other Peptides. To broaden our understanding of convergence times beyond the single EK peptide system, we performed REMD folding simulations of 13 additional peptides of 11−25 residues, as shown in Figures 6−10. Here, we ran only one REMD simulation for each peptide, not pursuing multiple runs owing to the size of the data set. For the same reason, we did not perform follow-up runs with an increased number of swap attempts per cycle. However, encouraged by the results we found for the EK peptide, we monitored the dominant cluster population during the progression of each simulation as an indicator of convergence, which is computed in 2 ns sliding windows.

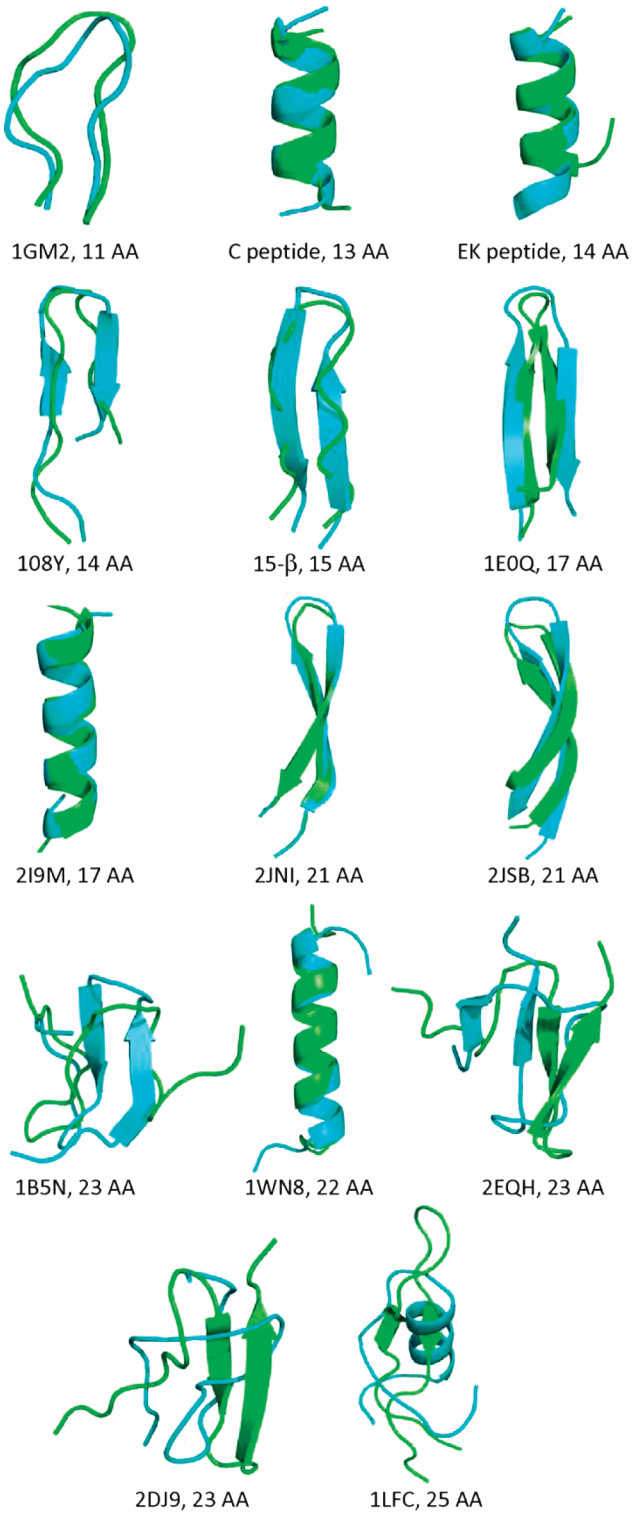

Figure 6. Native (green) versus simulated (blue) structures for peptides folded in this work. For the simulation structures, the dominant conformer from a clustering analysis of the final 10 ns of the simulation run is shown. The number of amino acids is also shown beneath each structure.

Each simulation continued until it had been deemed to reach equilibrium or appeared stable in the native structure, as assessed by the time progression of rmsd. In examining the results of these simulations, our main interests are the degree to which the top cluster population serves as a signal of

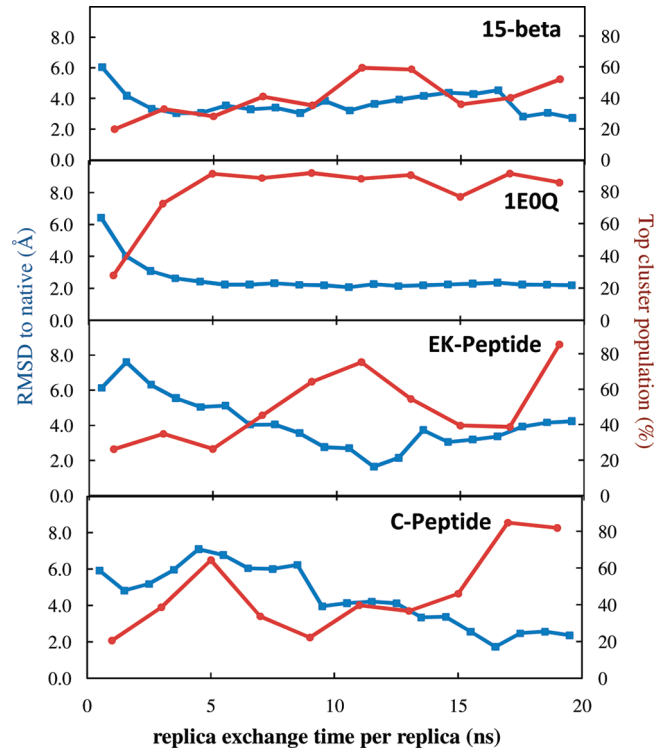

Figure 7. Rmsd and dominant cluster population versus time. Results are shown for the four fastest-folding peptides, whose simulation length was taken to 20 ns. All systems were started from an extended structure.

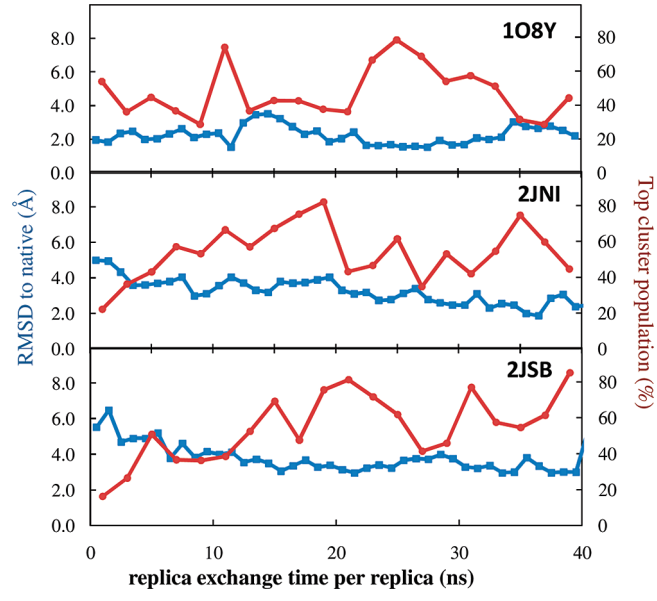

Figure 8. Rmsd and dominant cluster population versus time. Results are shown for peptides whose simulation length was taken to 40 ns. All systems were started from an extended structure.

convergence and the typical time scales involved in simulating to equilibrium this broad sampling of α and β peptides.

Figure 6 shows the top cluster structure from the final 5 ns of each simulation as compared to native, and Figures 7−10 give the rmsd as a function of time for each of the peptides, averaged over 1 ns windows. Of the 14 peptides simulated, 10 ultimately reach average RMSDs that are 4 Å

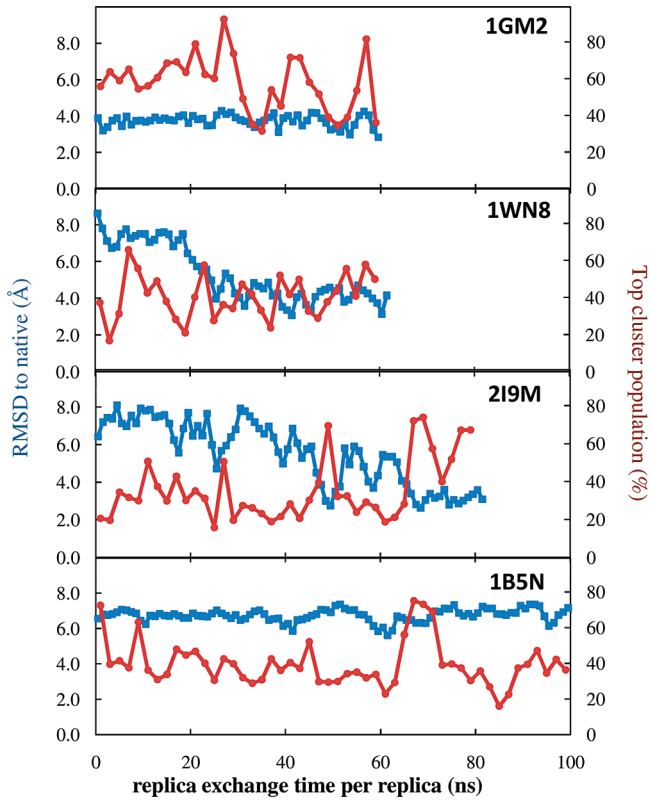

Figure 9. Rmsd and dominant cluster population versus time. Results are shown for peptides whose simulation length was taken to 60−100 ns. All systems were started from an extended structure.

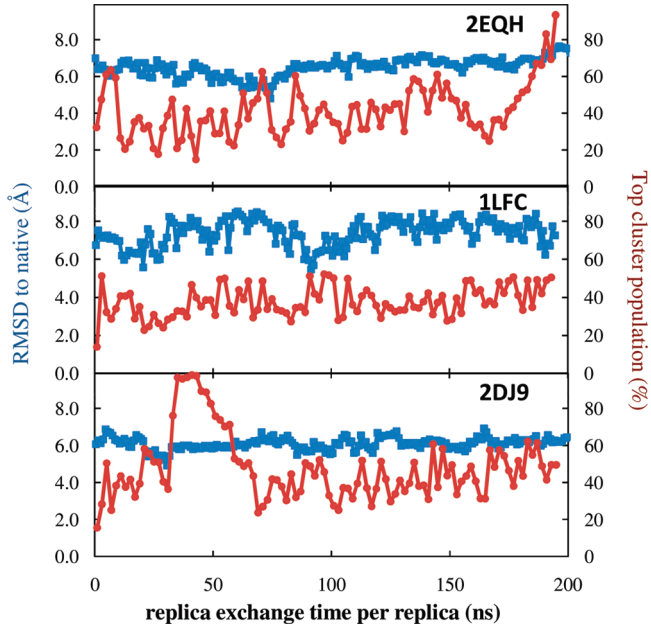

Figure 10. Rmsd and dominant cluster population versus time. Results are shown for peptides whose simulation length was taken to 200 ns. All systems were started from an extended structure.

or less, values that might be considered reasonable for convergence to the native basin given the conformational fluctuations that these short peptides are likely to experience. However, of these 10, there exists a variety of different folding behaviors. The majority of the peptides fold relatively

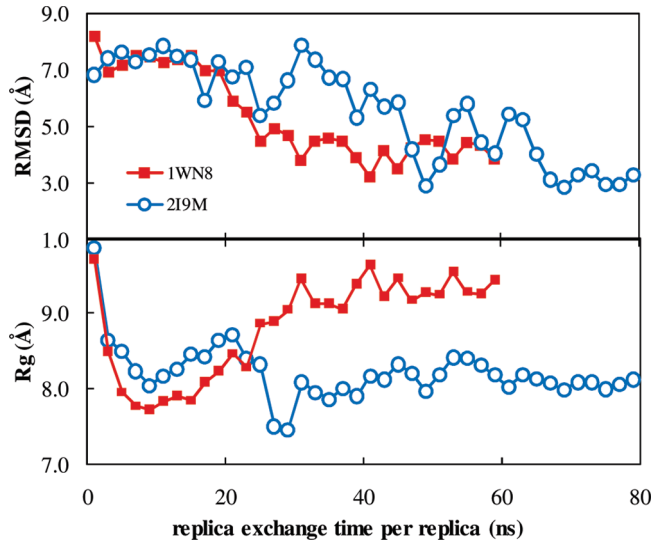

Figure 11. Rmsd and radius of gyration versus time for slow-folding helices. The runs for 1WN8 and 2I9M were terminated at 60 and 80 ns, respectively. Both systems were started from an extended structure.

fast in under 20 ns: the hairpin peptides 15-β, 1E0Q, 1O8Y, 2JNI, 2JSB, and the helical EK and C peptides. It is important to note that three of the hairpin peptides contained high contact order disulfide bonds, which undoubtedly provide a conformational restraint that funnels these simulations toward the native state. Still, 15-β, 1E0Q, and 1O8Y are the fastest folders, converging directly to the native basin in under 5 ns of simulation time, and the former two contain no disulfides at all.

The small 11-residue peptide 1GM2 also appears to converge quickly, its rmsd nearly constant with time from the start of the simulation. Yet, it exhibits a rather high average rmsd from the native structure, fluctuating around a value of 4 Å. The NMR structure for this peptide does contain several conformers, but these at most differ from each other by 1.3 Å, and therefore, the level of structural fluctuations seen in the REMD runs does not reflect the diversity in the PDB data. Still, the top cluster conformation from these runs in Figure 5 is very close to the native structure, a sign that the force field does favor the true native basin, but that significant fluctuations away from this structure raise the time-averaged rmsd. Moreover, it is not unreasonable that a short peptide of this size would exhibit marginal stability as compared to larger ones.

Two helical peptides take appreciably longer to fold. The 22-residue 1WN8 peptide begins to converge toward its native basin just after 20 ns, while the 17-residue 2I9M requires almost 60 ns of REMD simulation. The immediate question arises as to whether or not the slow-folding behavior of these two peptides is due to a β propensity in the backbone secondary structure preferences of the AMBER ff96 force field. In both cases, there is evidence that more specific driving forces play a role. In the case of 1WN8, it appears that a collapse to a dense and broken helix structure, driven by hydrophobic interactions, delays the formation of the full helix. Figure 11 compares the rmsd with the radius of gyration (R_g) versus time; the 1WN8 peptide experiences

an early sharp decrease in R_g that leads to the formation of a compact structure with a few helical turns. Eventually these turns coalesce, and the peptide proceeds to form the full helix with much higher R_g. While helical proteins have been shown to be fast folders,[67] one distinction here may be the artificial kinetics that REMD and implicit solvation impose.

On the other hand, the slow-folding behavior of the 2I9M helical peptide does not appear to be caused by the drive to a compact shape, as its R_g reaches an equilibrium value well before convergence is evident in the rmsd. We, therefore, clustered snapshots from the intermediate portion of the trajectory, 30−40 ns, and found that the dominant structure was a well-formed β-hairpin. On first look, such evidence might suggest β-propensity in the force field. A closer investigation, however, revealed that ion-pairing interactions between Lys-8, Arg-9, and Glu-12 and Glu-4 and Lys-16 permitted the formation of strong salt bridges that stabilized the turn and terminal hairpin regions well, respectively. Thus, one possible explanation for 2I9Ms slow folding might be the presence of kinetic trap conformations dominated by salt bridge interactions; salt bridges have been noted to affect folding in a number of previous studies.[46−48,52]

The simulations failed to fold four peptides, 1B5N, 2EQH, 1LFC, and 2DJ9. The average rmsd for these four peptides remains near a high constant value for nearly the entire run of 100−200 ns simulation time. All four peptides are long (23 or more residues), contain regions of β structure, have a single high contact order disulfide bond, and have a more complicated structure than that of the simple secondary structure elements of the other peptides. Given the success of the simulations with the other hairpin systems, these results are perhaps a bit surprising. One possibility is that, with the longer length of these peptides, the simulations simply require even more REMD time than currently performed. To assess this possibility, we performed for each a separate "native stability" REMD simulation, in which the initial conformation in all replicas was set to the native structure. Figure 12 shows the rmsd versus time for these runs, where it is clear that, in under 10 ns of simulation time, the system quickly loses the native structure. Therefore, we conclude that the native structure of these four peptides is not stable in the force field used.

Several factors can contribute to the failure of the folding and native stability simulations for these four peptides. One possibility is that the force field may not adequately stabilize larger, compact protein structures that have better-defined hydrophobic cores. This may result from inadequacies in the GBSA solvation model, but may also stem from the fact that a number of terms in the intramolecular force field have been parametrized to calculations on small molecule fragments and folding simulations of short peptides with single secondary structure elements. On the other hand, earlier work by us and others has shown that this force field can reach native-like structures in mechanism-based folding algorithms[51,53] and can stabilize native folds relative to decoys in "consensus" REMD simulations.[53] Another possibility for the failures is that these peptides might be only marginally

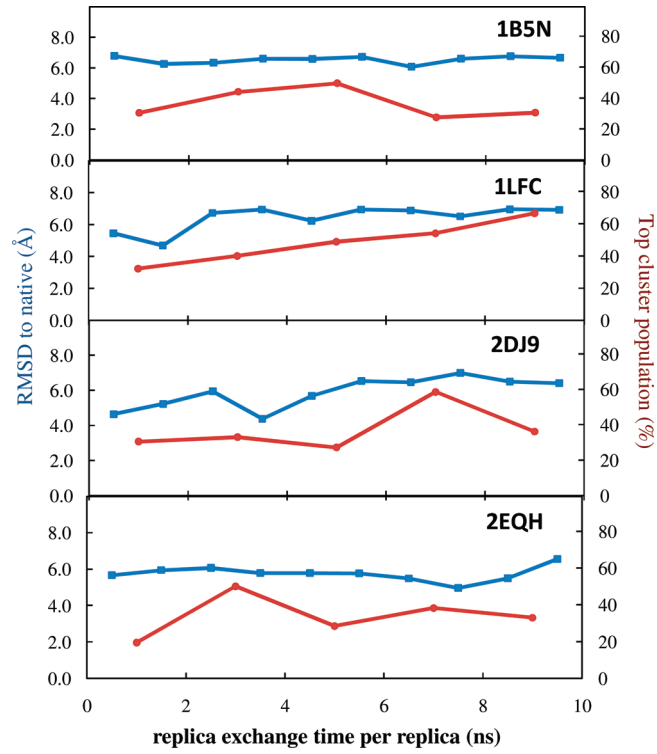

Figure 12. Rmsd and dominant cluster population versus time for "native stability" tests. Shown are the results for peptides that, starting from extended structures, were not able to be folded during 200 ns REMD runs. Here, all systems were started from native PDB structures. The results for each show a loss of structure after less than 2 ns worth of time.

stable in solution at standard conditions. All four have a weak secondary structure content, and all but 1B5N have multiple NMR models that indicate structural diversity in the terminal regions. Interestingly, our simulations of these locate structures with higher secondary structure content including both α and β motifs.

Throughout all of the peptide folding simulations, the dominant cluster population appears to be a fairly consistent predictor of convergence. In eight of the peptide runs (1E0Q, EK, C, 18OY, 2JNI, 2JSB, 1GM2, and 2I9M), a top cluster population of ∼70% or more is an indicator of convergence to an ensemble of structures with an average rmsd typical of the equilibrium and folded average rmsd. There are, however, several cases that do not follow this rule. For the 15-β peptide, the criterion must be reduced to the ∼60% level, as this peptide never exceeds this threshold population but does correctly fold. For 1WN8, the cluster population reaches 60% at several points during the simulation but does not appear to signal low rmsd values; longer simulation times may show higher populations in this peptide, although we did not pursue them. However, the top cluster criterion also signals some false positives: in three of the peptides that did not fold correctly, it reaches very high values, greater than 90%, at several points during these runs. Part of this behavior may be due to some of the problems with these peptides mentioned earlier, such as their low secondary structure content. Still, this finding does suggest that the

cluster population, while useful, is probably not a foolproof indicator of convergence in all cases.

Conclusions

In this study, we assessed the convergence properties of the REMD method for a number of peptide systems displaying both α and β secondary structure motifs, using an AMBER force field with a GBSA solvation model. We performed a detailed study of the 14-residue helical EK peptide,[55] conducting five separate 100 ns REMD runs each, starting with an initial unfolded conformation and different velocity set. We then folded 13 additional peptides taken from the PDB and earlier simulation studies. Folding behavior was assessed using backbone rmsd distances from experimental (if available) or idealized (e.g., helical) structures, and a clustering analysis was used to extract dominant conformational states.

Our results suggest several basic features of REMD simulations that are essential to consider in folding studies and, in particular, when using these methods to evaluate force field performance:

Heterogeneity. We find that run-to-run heterogeneity can be quite large. Our multirun study of the EK peptide finds slow-, fast-, and misfolding trajectories that span convergence times from under 10 ns to 30−40 ns. In all runs, folding events are found to occur across a spectrum of temperatures, spanning the lowest replica temperature to near the folding temperature. Moreover, the five different trials for this peptide do not appear to look indistinguishable until reaching a time of about 60 ns. Such heterogeneity is not unexpected given its existence in even fast-folders simulated with molecular dynamics runs,[50] but it has not yet become common practice to perform multiple trials in REMD folding simulations perhaps due to computational expense. While the degree of heterogeneity observed is likely to be sequence dependent, the possibility of its existence underscores the importance of long and/or multiple simulation trials to achieve a more comprehensive and definitive analysis of folding behavior. One attractive possibility for reducing run-to-run variances is to increase the number of swap attempts performed during REMD cycles; in our study of the EK peptide, we find this approach removes slow-folding trajectories and reduces mean convergence times by a factor of almost three.

Fluctuations. Our simulations demonstrate the presence of significant fluctuations in the converged portions of the folding runs. The average rmsd found for the peptides that do converge to equilibrium and fold to a correct structure varies between 2−4 Å. Typically the dominant cluster structure exhibits an rmsd that is lower than the average, and the nondominant clusters represent fluctuations around it. The average rmsd values that we find are consistent with earlier simulation studies.[12,17,27,30,33,36] Importantly, these observations reinforce a picture of peptides existing not in static, single structures but in conformational ensemble whose fluctuations might play an important role in their functions in living systems.

Convergence Times. We find that convergence times can be quite large, at least, when compared to the time scales reached in many earlier efforts. The majority of our peptide folding simulations converge in under 20 ns of REMD time (per replica), but a few systems, notably all helical, require up to 60 ns to reach equilibrium. The factors contributing to these slow-folding times appear to be somewhat distinct for each case, and strong salt bridges seem to contribute to slow-folding in only one of these systems. Importantly, we find a fairly reliable approach for signaling simulation convergence in monitoring the time progression of the population of the largest conformational cluster, which is in agreement with earlier efforts.[34−36] Importantly, this population metric is independent of the native structure and, thus, suitable for blind, de novo folding studies. For the EK peptide, the dominant cluster population is able to delineate convergence times across the slow- and fast-folding trajectories. For the majority of the other peptides we examined, cluster populations of ∼60−70% correspond well with convergence to the native basin. However, this metric is not foolproof; in a few cases, notably those peptides that we were not able to successfully fold, the cluster population gives some false positives for convergence. In future work, the combination of cluster population with other indicators, such as backbone entropies or correlation times, may improve the robustness of this convergence signal.

While these results may be specific to the particular force field that we use, they nevertheless demonstrate the possibility of the kinds of behavior that may accompany any force field and, thus, emphasize the need for addressing convergence issues in force field assessment. Even so, we do find that the combination of AMBER ff96[62] with OGB implicit solvation[63] is able to correctly fold 10 of the 14 peptides studied, a pool including four α-helical and six β-hairpin structures. Its failure to fold the other four peptides may be related to their low secondary structure content, more complicated tertiary arrangement, and unstructured and flexible regions. Overall, this work suggests that the systematic application of long runs, multiple trials, convergence indicators, and diverse peptide databases may offer new perspectives that can improve force field development efforts.

Acknowledgment. We thank the authors in ref 56 for providing the structure of the 15-β peptide. M.S.S. gratefully acknowledges the support of a Camille & Henry Dreyfus Foundation New Faculty Award and a UCSB − Santa Barbara Cottage Hospital Special Research Award.

References

(1) 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*, 6127–6129.

(2) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203.

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

(4) Hansmann, U. H. E.; Okamoto, Y. Tertiary Structure Prediction of C-Peptide of Ribonuclease A by Multicanonical Algorithm. *J. Phys. Chem. B* **1998**, *102*, 653–656.

(5) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: molecular dynam-

ics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, *284*, 835–848.

(6) Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.

(7) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740–744.

(8) Ono, S.; Nakajima, N.; Higo, J.; Nakamura, H. Peptide free-energy profile is strongly dependent on the force field: Comparison of C96 and AMBER95. *J. Comput. Chem.* **2000**, *21*, 748–762.

(9) Ferrara, P.; Caflisch, A. Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780–10785.

(10) Jang, S.; Shin, S.; Pak, Y. Molecular Dynamics Study of Peptides in Implicit Water: Ab Initio Folding of Beta-Hairpin, Beta-Sheet, and Beta-Beta-Alpha-motif. *J. Am. Chem. Soc.* **2002**, *124*, 4976–4977.

(11) 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*, 21–31.

(12) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.

(13) Zagrovic, B.; Sorin, E. J.; Pande, V. Beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **2001**, *313*, 151–169.

(14) Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **2002**, *323*, 927–937.

(15) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102–106.

(16) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. Peptide folding simulations. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(17) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **2001**, *42*, 345–54.

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

(19) Yoda, T.; Sugita, Y.; Okamoto, Y. Secondary structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem. Phys.* **2004**, *307*, 269–283.

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

(21) Yoda, T.; Sugita, Y.; Okamoto, Y. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett.* **2004**, *386*, 460–467.

(22) Anderson, R. E.; Pande, V. S.; Radke, C. J. Dynamic lattice Monte Carlo simulation of a model protein at an oil/water interface. *J. Chem. Phys.* **2000**, *112*, 9167–9185.

(23) Rao, F.; Caflisch, A. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.* **2003**, *119*, 4035–4042.

(24) Jang, S.; Kim, E.; Pak, Y. Free energy surfaces of miniproteins with a $\beta\beta$a motif: Replica exchange molecular dynamics simulation with an implicit solvation model. *Proteins* **2006**, *62*, 661–671.

(25) Kim, E.; Jang, S.; Pak, Y. Consistent free energy landscapes and thermodynamic properties of small proteins based on a single all-atom force field employing an implicit solvation. *J. Chem. Phys.* **2007**, *127*, 145104.

(26) Jang, S.; Kim, E.; Pak, Y. Direct folding simulation of alpha-helices and beta-hairpins based on a single all-atom force field with an implicit solvation model. *Proteins* **2007**, *66*, 53–60.

(27) Kim, E.; Jang, S.; Pak, Y. Direct folding studies of various a and β strands using replica exchange molecular dynamics simulation. *J. Chem. Phys.* **2008**, *128*, 175104.

(28) Kim, E.; Yang, C.; Jang, S.; Pak, Y. Free energy landscapes of a highly structured β-hairpin peptide and its single mutant. *J. Chem. Phys.* **2008**, *129*, 165104.

(29) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(30) Lei, H.; Wu, C.; Wang, Z. X.; Zhou, Y.; Duan, Y. Folding processes of the B domain of protein A to the native state observed in all-atom ab initio folding simulations. *J. Chem. Phys.* **2008**, *128*, 235105.

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

(32) Periole, X.; Mark, A. E. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.* **2007**, *126*, 014903.

(33) Lei, H.; Wu, C.; Liu, H.; Duan, Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4925–4930.

(34) Daura, X.; van Gunsteren, W. F.; Mark, A. E. Folding-unfolding thermodynamics of a-heptapeptide from equilibrium simulations. *Proteins* **1999**, *34*, 269–280.

(35) Smith, L. J.; Daura, X.; van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simulations. *Proteins* **2002**, *48*, 487–496.

(36) Lyman, E.; Zuckerman, D. M. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.* **2006**, *91*, 164–172.

(37) Abraham, M. J.; Gready, J. E. Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1119–1128.

(38) Gnanakaran, S.; Garcia, A. E. Validation of an all-atom protein force field: from dipeptides to larger peptides. *J. Phys. Chem. B* **2003**, *107*, 12555–12557.

(39) Mu, Y.; Kosov, D. S.; Stock, G. Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. *J. Phys. Chem. B* **2003**, *107*, 5064–5073.

(40) Best, R. B.; Buchete, N. V.; Hummer, G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* **2008**, *95*, 7–9.

(41) Sorin, E. J.; Pande, V. S. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472–2493.

(42) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.

(43) Hu, H.; Elstner, M.; Hermans, J. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins* **2003**, *50*, 451–463.

(44) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(45) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777–12782.

(46) Zhou, R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* **2003**, *53*, 148–161.

(47) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins* **2004**, *56*, 310–321.

(48) 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.

(49) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. Secondary structure bias in generalized Born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *J. Phys. Chem. B* **2007**, *111*, 1846–1857.

(50) Ensign, D. L.; Kasson, P. M.; Pande, V. S. Heterogeneity Even at the Speed Limit of Folding: Large-scale Molecular Dynamics Study of a Fast-folding Variant of the Villin Headpiece. *J. Mol. Biol.* **2007**, *374*, 806–816.

(51) Ozkan, S. B.; Wu, G. H. A.; Chodera, J. D.; Dill, K. A. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11987–11992.

(52) Shell, M. S.; Ritterson, R.; Dill, K. A. A test on peptide stability of AMBER force fields with implicit solvation. *J. Phys. Chem. B* **2008**, *112*, 6878–6886.

(53) Shell, M. S.; Ozkan, S. B.; Voelz, V. A.; Wu, G. H. A.; Dill, K. Blind test of physics-based prediction of protein structures. *Biophys. J.* **2009**, *96*, 917–924.

(54) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247*, 536–540.

(55) Scholtz, J. M.; Barrick, D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. Urea Unfolding of Peptide Helices as a Model for Interpreting Protein Unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 185–189.

(56) Santiveri, C. M.; Pantoja-Uceda, D.; Rico, M.; Jimenez, M. A. Beta-hairpin formation in aqueous solution and in the presence of trifluoroethanol: a (1)H and (13)C nuclear magnetic resonance conformational study of designed peptides. *Biopolymers* **2005**, *79*, 150–162.

(57) Bierzynski, A.; Kim, P. S.; Baldwin, R. L. A Salt Bridge Stabilizes the Helix Formed by Isolated C-Peptide of RNase A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2470–2474.

(58) Pitera, J. W.; Haque, I.; Swope, W. C. Absence of reptation in the high-temperature folding of the trpzip2 β-hairpin peptide. *J. Chem. Phys.* **2006**, *124*, 141102.

(59) Ho, B. K.; Dill, K. A. Folding very short peptides using molecular dynamics. *PLoS Comput. Biol.* **2006**, *2*, e27.

(60) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384.

(61) Mehlig, B.; Heermann, D. W.; Forrest, B. M. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 679.

(62) Kollman, P. A.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. *Comput. Simul. Biomol. Syst.* **1997**, *3*, 83–96.

(63) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3261–3429.

(64) García, A. E.; Sanbonmatsu, K. Y. alpha-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(65) Gnanakaran, S.; Garcia, A. E. Folding of a highly conserved diverging turn motif from the SH3 domain. *Biophys. J.* **2003**, *84*, 1548–1562.

(66) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

(67) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998**, *277*, 985–994.

CT900119N