

PACE Force Field for Protein Simulations. 2. Folding Simulations of Peptides

Wei Han,[†] Cheuk-Kin Wan,[†] and Yun-Dong Wu^{*,†,‡,§}

Department of Chemistry, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, China, School of Chemical Biology and Biotechnology, Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen, China, and College of Chemistry, Peking University, Beijing, China

Received June 8, 2010

Abstract: We present the application of our recently developed PACE force field to the folding of peptides. These peptides include α -helical (AK17 and Fs), β -sheet (GB1m2 and Trpzip2), and mixed helical/coil (Trp-cage) peptides. With replica exchange molecular dynamics (REMD), our force field can fold the five peptides into their native structures while maintaining their stabilities reasonably well. Our force field is also able to capture important thermodynamic features of the five peptides that have been observed in previous experimental and computational studies, such as different preferences for a helix–turn–helix topology for AK17 and Fs, the relative contribution of four hydrophobic side chains of GB1p to the stability of β -hairpin, and the distinct role of a hydrogen bond involving Trp-H_e and a D9/R16 salt bridge in stabilizing the Trp-cage native structure. Furthermore, multiple folding and unfolding events are observed in our microsecond-long normal MD simulations of AK17, Trpzip2, and Trp-cage. These simulations provide mechanistic information such as a “zip-out” pathway of the folding mechanism of Trpzip2 and the folding times of AK17 and Trp-cage, which are estimated to be about 51 ± 43 ns and 270 ± 110 ns, respectively. A 600 ns simulation of the peptides can be completed within one day. These features of our force field are potentially applicable to the study of thermodynamics and kinetics of real protein systems.

Introduction

The successful application of molecular dynamics (MD) simulations to structural and dynamic studies of proteins relies on the efficiency of the sampling protocol and the quality of the underlying force fields.^{1,2} With the rapid growth in computing power, advanced computing techniques³ and accelerating sampling methods such as multicanonical simulations,⁴ replica exchange molecular dynamics (REMD),⁵ and metadynamics,⁶ computer simulations have become far more powerful than those of even a few years ago.^{7–9} Despite this progress, all-atom force fields with

explicit treatment of solvents, which are arguably the most accurate, are only capable of attaining converged sampling for small peptides with not more than 20 amino acids at considerable computational expense.^{9–11}

To further enhance sampling, it is necessary to reduce the number of degrees of freedom. The key challenge here is how to simplify models with a minimal loss of accuracy. Several approaches have been developed. For example, an explicit solvent model can be replaced by an implicit solvent model such as the generalized Born model with surface area (GB/SA).^{12–14} In conjunction with all-atom protein models, GB/SA has been accepted and used widely for its improved efficacy.¹ Very recently, the combination of the AMBERff96 all-atom force field and the GB model by Onufriev et al.¹⁴ has made a breakthrough in folding a millisecond slow folder NTL9(1–39),⁷ with the help of large-scale distributed

* Corresponding author e-mail: chydwu@ust.hk.

† The Hong Kong University of Science & Technology.

‡ Peking University Shenzhen Graduate School.

§ Peking University.

Table 1. Sequences and Experimental Stabilities of the Five Peptides Studied in This Work and Previous Representative Computational Studies of These Peptides

peptide name	sequence ^a	stability at 300 K	refs	used to optimize parameters? ^b
AK17	Ac-(AAKAA) ₃ GY-Nme	~30–35% ^c	42, 46	overall helical content
Fs	Ac-A ₅ (AARAA) ₃ -Nme	~50–55%; $T_m = 303\text{ K}^c$	43–45, 62, 71, 84	no
GB1m2	GEWTYNPATGKFTVTE	~70–75%; $T_m = 320\text{ K}^d$	68, 71	overall β -sheet content
Trpzip2	SWTWENCKWTWK	>90%; $T_m = 345\text{ K}^d$	62, 68, 69, 98, 99	no
Trp-cage	NLYIQWLKDGGPSSGRPPPS	~70%; $T_m = 317\text{ K}^d$	9, 52–58, 68, 70, 71	no

^a The underlined residues are in helical states; the bold residues are in sheet states; the italic residues are in loops, turns, or coils. ^b If yes, this column indicates which propensities of the peptides are used for parametrization. ^c Based on CD measurement.^{79,80} ^d Based on NMR measurement.^{33,35,36}

computations and GPU coding.³ Besides the implicit model, another approach is to use a coarse-grained (CG) protein representation,^{15,16} which is coupled with either an implicit^{17–19} or a CG solvent model.^{20–22}

Combined with the fast sampling method REMD, the simplified model has reached the point where converged sampling for small peptides and mini-proteins only costs moderate computation time. Thus, a comparison with experimental observations can be readily made. This comparison is a critical test of the quality of force fields even for all-atom models with explicit solvents, as most of the models are derived from small molecules or dipeptide models.^{23–26} On the experimental side, there are a number of designed peptides that fold at the microsecond time scale.²⁷ These peptides possess α -helices,^{28–30} β -hairpins,^{31–35} or mixed topologies,^{36–41} and tertiary structures can be found in some of them.^{36–41} As a result, they are perfect targets in extensive computational studies.^{9,19,40,42–59}

The comparisons reveal that it is not trivial for force fields to achieve a balance between α -helical and β -sheet structures. Okamoto and co-workers have shown that different all-atom force fields with explicit solvent models have quite different preferences for helical and extended structures.^{60,61} Shell et al. performed extensive REMD simulations of short helices and hairpins with four versions of AMBER force fields and three sets of parameters for the GB/SA implicit solvent, which were proposed previously.⁶² They reported that only the AMBERff96 coupled with the GB/SA parameters proposed by Onufriev et al.¹⁴ could attain the balance. A CG model with implicit solvent has also been found to be problematic in transferability when used to predict α , β , and mixed structures.⁶³ The peptide simulations clearly demonstrate that current force fields need further optimization.^{64–67} A good balance for a number of small peptides has been achieved in several encouraging cases such as a modified version of the CHARMM22/CMAP with the GB solvent by Chen et al.,⁶⁸ the OPLS-AA with the GB model by Ulmschneider and Jorgensen,⁶⁹ the CG OPEP force fields by Chebaro et al.,⁷⁰ and an all-atom model with implicit solvent by Irback and Mohanty.⁷¹

In our preceding paper,⁷² we presented the full parametrization of a united-atom protein model, namely, the PACE (Protein in Atomistic details coupled with Coarse-grained Environment) force field, for all 20 amino acids, which is coupled with a CG solvent model.^{73–76} We have shown that the force field is capable of maintaining the native structures of several proteins in MD simulations. It is thus critical to

evaluate the applicability of the force field in folding simulations of peptides, and in particular to examine whether the force field attains a balance between the α -helix and β -sheet. In this paper, we report the results of the folding simulations of five small peptides using REMD techniques. As listed in Table 1, these peptides include (1) designed α -helical peptides such as AK17²⁹ and Fs⁷⁷ peptides; (2) β -hairpin peptides such as GB1m2,³³ a mutant of the N-terminal hairpin in the protein G B1 domain that has much faster folding kinetics than the wide type, and Trpzip2,³⁵ and (3) the α -helical/coil Trp-cage³⁶ that possesses a tertiary topology. These peptides have been extensively studied in MD simulations and used to validate parameters of force fields (Table 1). As normal MD can also provide kinetic information about folding, we have also carried out normal MD folding simulations for AK17, Trpzip2, and Trp-cage. In addition, possible further improvements on this model will also be discussed.

Model and Methods

All of the simulations were performed with the GROMACS software package (version 3.3.1).⁷⁸ The PACE force field, presented in the preceding paper, was used for peptides.^{21,22,72} Water was modeled by the CG solvent model of Marrink et al.⁷³ All of the peptides were capped with an acetyl group at N-terminals and N-methylamide at C-terminals. All of the starting conformations were full helices. The peptide was placed into a dodecahedron box. The minimum distance between the peptide and the edges of the box was about 1.3 nm. About 1100–1300 CG solvent particles, depending on the size of the peptide, were filled in the box. The whole system was first optimized with the steepest descent method by 5000 steps. A 120 ps pre-equilibrium simulation was carried out on the optimized system at $T = 300\text{ K}$ and $P = 1\text{ atm}$ with a time step of 6 fs. A 6 ns NVT denaturing simulation was then performed at $T = 700\text{ K}$. Snapshots after 2 ns of the denaturing simulation were randomly chosen, and 120 ps simulations of these snapshots were performed at $T = 300\text{ K}$ and $P = 1\text{ atm}$ to prepare the starting conformations for the production simulation. The time step of the denaturing simulation is 4 fs due to the high temperature.

Our REMD simulations were composed of 16 replicas. Their temperatures ranged from 300 to 430 K. The pressure was kept at 1 atm. A time step of 10 fs was used. The mass of peptides was tripled to avoid the instability of the

simulations at high temperatures. It will be shown later that this has little effect on the thermodynamic properties of the simulated systems. An attempt to exchange configurations between replicas was made every 2 ps. The average ratio of a successful exchange was about 15%. With a cluster of eight two-way 2.66 GHz CPUs, the wall time for 300 ns ($\times 16$ replicas) of REMD simulation was about 30 h.

Following Garcia and Sanbonmatsu,⁴⁴ a residue was considered as helical only if this residue and its two neighboring residues had their (ϕ , Ψ) within $(-60^\circ \pm 30^\circ, -47^\circ \pm 30^\circ)$. The helical probability (or fraction) is calculated as the ratio of the number of helical residues to the total residue number minus two, which accounts for two terminal residues that can never be helical. A hydrogen bond was considered formed only if the donor/acceptor distance was shorter than 0.35 nm and the donor/hydrogen/acceptor angle was larger than 120° .

Results and Discussions

Simulations of Helical Peptides. The total helical content of the AK17 peptide at 300 K is $\sim 41\%$, which is averaged over the last 70 ns of 100 ns REMD simulations. Since this quantity has been used to fit our force field parameter,⁷² as expected, it is consistent with $\sim 30\%$ of the helical content of this peptide by CD measurements.⁷⁹ It is natural to immediately wonder if the optimized parameters can be transferred to other helical peptides. We therefore also carried out the REMD simulations of a longer helical peptide [Ac–A₅(AAARA)₃A–NMe], known as the Fs peptide,^{30,77} which has been explored experimentally.^{80,81} The CD-based measurement gave a melting temperature T_m of 303 K for this peptide.⁸⁰ The average helical content in our simulation (the last 50 ns of 80 ns REMD simulation) is $\sim 43\%$, in agreement with experimental results.

In addition to the total helical content, the residual helical contents of the AK17 and Fs peptides at 300 K were investigated (Figure 1). Their residual helical contents share two common features: (1) The residues in the N-terminus have a higher helical content than the C-terminal residues. In AK17, residues 2 and 3 at the N-terminus have helical contents of $\sim 40\%$ and $\sim 60\%$, respectively, while residues 14–16 at the C-terminus have helical contents of $\sim 0\text{--}20\%$. In Fs, residues 2 and 13 are $\sim 40\%$ and $\sim 65\%$ helical, respectively, but residues 19 and 20 are $\sim 25\%$ and $\sim 10\%$ helical, respectively. The asymmetry of the helical content distribution has been well characterized by previous NMR studies of polyalanine-based peptides.^{82,83} (2) The residual contents in the middle of the helices are slightly lower than the residual contents closer to the terminals. In AK17, the middle residues 7–9 are $\sim 40\%$ helical, but the highest residual helical content exceeds 60%. In Fs, the helical contents of residues 10 and 11 are also $\sim 40\%$, lower than those of residues 3–6 and 15–17. This feature has also been observed in the previous MD simulation studies of the A21 and Fs peptides.^{43,45}

The Fs peptide has been the subject of numerous MD studies.^{43–45,62,71} Using the AMBER force field with the GB solvent, Zhang et al. performed multiple 100 ns MD

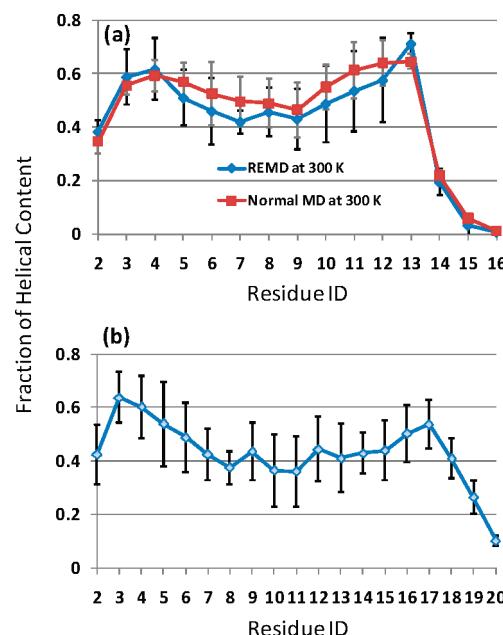


Figure 1. Residual helical contents of the AK17 (a) and Fs 17 (b) peptides at 300 K. The error bars were estimated from block averages, with block sizes of 10 and 500 ns for REMD and normal MD simulations, respectively.

simulations and found that a helix–turn–helix structure is predominant for Fs at 300 K.⁴³ In a seminal work, Garcia and Sanbonmatsu⁴⁴ applied the REMD techniques so that equilibrium sampling could be achieved. They found that the AMBER-ff94 force field with explicit solvent significantly overestimates the T_m of Fs ($T_m \sim 400$ K). This can be corrected by ignoring all backbone torsion terms in this force field. Using the modified force field, they suggested that Fs' preference for the helix–turn–helix structure in the GB simulations is likely an artifact of the implicit solvent.⁸⁴ Sorin and Pande⁴⁵ proposed another modified version of the AMBER force field (parm99 ϕ) with explicit solvent that not only renders correct thermodynamics of this peptide but also reproduces the non- α conformations and helix–coil kinetics better than Garcia's modified version. Interestingly, their results showed that the population of the helix–turn–helix is considerable.

To reveal the structural features of our Fs model in more detail, we clustered the structures sampled from the 30th to 100th nanosecond of the simulation at 300 K according to the root-mean-square deviation (RMSD) of backbone atoms with a cutoff of 0.25 nm. A total of 125 clusters were identified, and the top 10 clusters which account for $>50\%$ of the total population are shown in Figure 2a. Clearly, various helical topologies can be sampled with our force fields. Following Zhang et al.,⁴³ we defined seven helical topologies, tagged them on each cluster, and then counted the overall populations of each topology. The topologies include (1) full helix (F–H), (2) one-helix (1-H), which has only one helical fragment, (3) helix–turn–helix (HTH), which has two helical fragments and an intervening loop/turn in the middle of the peptide, (4) N-helix–turn–helix (N-HTH), which has a helix–turn–helix topology with the loop/turn closer to the N-terminus, (5) C-helix–turn–helix (C-HTH) with the loop/turn closer to the C-terminus, (6)

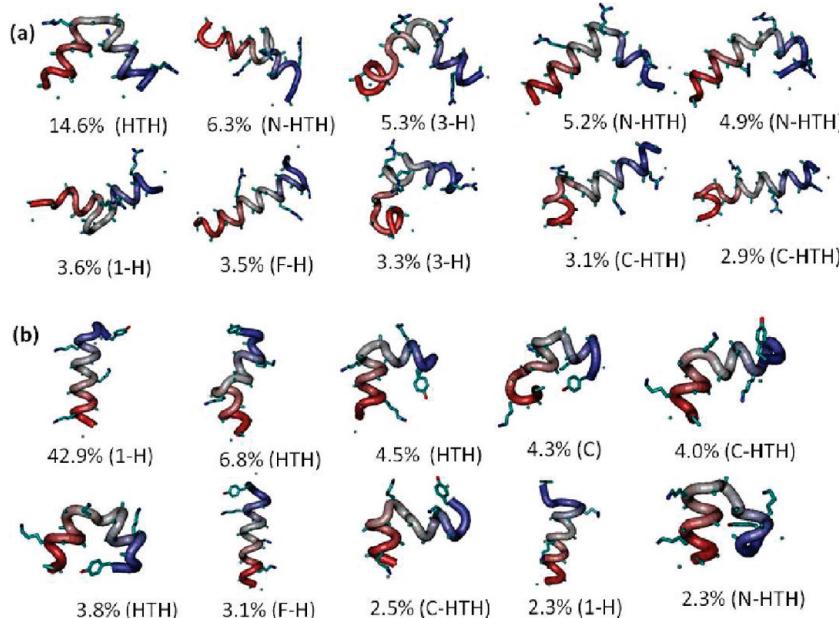


Figure 2. Representative structures of the most populated clusters of the Fs (a) and AK17 (b) peptides.

Table 2. Probabilities of Different Helical Topologies Sampled in the Simulations of the Fs and AK17 Peptides

	Fs	AK17
full helix	3.5%	3.1%
one helix ^a	17.1%	55.0%
helix–turn–helix ^b	19.3%	15.7%
N-helix–turn–helix ^c	20.0%	6.8%
C-helix–turn–helix ^d	15.7%	7.4%
three-turn–helix ^e	19.9%	7.0%
full coil	4.5%	5.0%

^a A conformation with only one helical fragment. ^b Two helical fragments intervened by a loop/turn in the middle of the peptide.

^c A helix–turn–helix with the loop/turn closer to the N-terminus.

^d A helix–turn–helix with the loop/turn closer to the C-terminus.

^e A conformation with three helical fragments.

three-turn–helix (3-H), which has three helical fragments, and (7) full coil (C).

It can be seen from Table 2 that the structures with a two-helix topology are the most abundant in our Fs simulations. The overall population of HTH, N-HTH, and C-HTH is ~55%, comparable to the ~53% in implicit solvent simulations⁴³ and slightly larger than the ~42% in explicit solvent simulations.⁴⁵ The probability of a full helical structure is ~3%, smaller than the ~20% and the ~17% estimated from the implicit solvent⁴⁵ and explicit solvent models,⁴⁵ respectively. This is because, in our Fs model, the fraying ends prevent the formation of a perfect full helix. If the one-helix topology is also considered, the combined population reaches ~21%, which is much closer to the previous studies. The topological distribution here can also be cross-checked using the radius of gyration (Rg) of the conformational ensemble. Small-angle X-ray scattering (SAXS) measurements reported an Rg value of ~0.9 nm for the Fs.⁴⁵ Our calculated $\langle Rg \rangle$ is ~0.85 ± 0.07 nm, slightly smaller than but still in agreement with experimental results.

One may wonder if the preference to the HTH topologies is inherent to our model. We therefore conducted the same analysis for the AK17 peptide. The top 10 clusters and the

populations of each helical topology are shown in Figure 2b and Table 2, respectively. The most stable cluster forms almost all possible helical HBs (on average ~12 helical HBs) except for the C-terminal Gly. The backbone of the Gly residue, known as a helix breaker,⁸⁵ adopts an α_L conformation in the most stable cluster. Even though the AK17 peptide has only four fewer residues than the Fs peptide, the structures with only one helical fragment make up ~58% of the total population. On the other hand, the population of two-helix topologies decreases to ~30%. This is in line with an implicit solvent MD study by Chowdhury et al.⁴² where a similar peptide [Ac–YG(AAKAA)₂AKA–NH₂] was found to be ~60% full helix. Moreover, the calculated Rg of the AK17 is 0.78 ± 0.06 nm, close to the experimental value of 0.82 nm.⁴⁶ The topological analyses of the Fs and AK17 peptides demonstrate that our force field is able to capture the common topological features observed in previous implicit and explicit solvent models.

Besides the REMD simulations, a 3 μ s normal MD simulation was performed for the AK17 at 300 K. Our purpose was 2-fold. In the REMD simulation, a 10 fs time step was used, and the mass of peptides is tripled to maintain replicas at high temperatures. In the normal MD, a 6 fs time step was used without scaling of the mass of the peptide. This time step is well below the upper limit of a 7.5 fs time step for the conservation of the energy of a system when using dummy atoms in simulations.⁸⁶ Thus, any effect of the REMD setup on the thermodynamics properties can be seen from the comparison with normal MD results. Our second purpose was to examine how fast our model can sample the conformational space with a normal MD protocol. Unlike the REMD, a normal MD is able to provide information about the transition between states, which can be valuable for an assumption-free analysis of the kinetics of protein folding.^{7,8}

The helical content and the average Rg during the normal MD simulation are ~43% and 0.79 ± 0.06 nm, respectively,

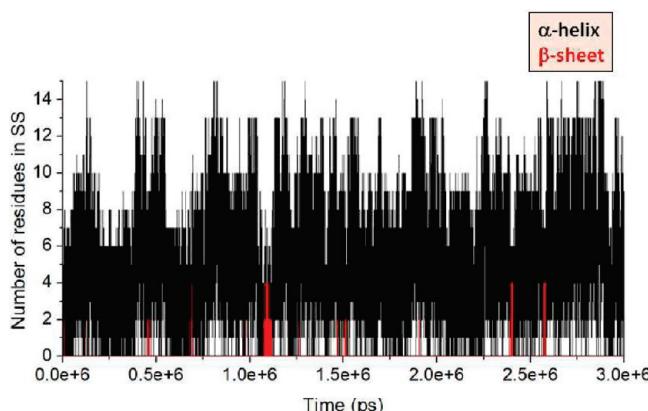


Figure 3. Change in the numbers of helical and sheet residues during a normal MD simulation of the AK17 peptide at 300 K.

which agree well with our REMD results (41% and 0.78 ± 0.06 nm, respectively). In addition, the residual helical contents of each amino acid of the REMD simulation also match well with those of the normal MD simulations (Figure 1a). These suggest that our REMD protocol has little effect on the thermodynamics.

Figure 3 shows the change in the numbers of residues in helices and sheets during the normal MD simulations. Multiple helix–coil and sheet–coil transitions are observed. Although a detailed kinetic analysis of the folding mechanism is out of the scope of this paper, we wanted to roughly estimate how fast the model travels in the conformational space. Following what we have done before,²¹ we tracked the dynamic change in the number of helical residues. After a coil is reached (zero helical residues), a folding event is thought to occur once a nearly full helix is formed (≥ 12 helical residues). An unfolding event is assumed as a transition from a nearly full helix to a coil. In the 3 μ s simulations, there are a total of 32 folding/unfolding events. The averaged times for folding (τ_f) and unfolding (τ_u) events are $\sim 51 \pm 43$ and $\sim 44 \pm 33$ ns, respectively, which are similar to our previous calculations.²¹ The relaxation time (τ) of helix–coil transition in our MD simulation is estimated to be ~ 24 ns according to $\tau = \tau_u \tau_f / (\tau_u + \tau_f)$. The kinetics of helix–coil transition here are approximately an order of magnitude faster than the helix–coil kinetics from experimental measurements ($\tau \sim 100$ –500 ns).^{81,80,87,88} Two reasons may account for the acceleration. One is the CG solvent model that we used. Marrink et al. have shown that coarse-graining on water can lead to an overall increase in the diffusion rate of solutes by 4–5 times.⁷³ The other factor may be our simplification to the protein model, leading to a smoother free energy surface for the helix–coil transition.

Simulations of β -Hairpin Peptides. Folding of the GB1m2 Peptide. With the same force field, we now turn to β -sheet peptides. The peptide GB1p (GEWTYDDATKTFVTE), the N-terminal fragment of the protein G B1 domain, has been known to possess a native-like β -hairpin structure in aqueous solution.³² Although a Trp fluorescence study reported that this peptide is 80% folded at 297 K and has a folding time of $\sim 6 \mu$ s,⁸⁹ recent NMR data suggested that its folded population is only $\sim 30\%$ at 298 K³³ and its

folding time is 17–20 μ s,³⁴ slower than previously estimated. Several mutants of GB1p have been derived to increase the stability and folding speed of the β -hairpin.³³ For instance, a loop mutant, GB1m2 (GEWTYNPATGKFTVTE), can have $\sim 74\%$ folded structures at 298 K, and its folding time decreases to $\sim 5 \mu$ s.^{33,34}

Because of their well-characterized kinetics and thermodynamics, the GB1p series have been investigated in numerous computational studies.^{47–51,62,68,70,71,90} They have been used to test the quality of force fields. For example, the AMBER ff99SB and ff99 sets of parameters with the GB solvent fold the GB1p into helical structures instead of β -hairpins.⁶² On the other hand, several force fields such as OPLS-AA and the AMBER ff03 and ff99ci sets overestimate the stability of GB1p considerably.⁹⁰ The calculated T_m 's are 60–80 K above the experimental T_m values (293 K). Interestingly, using their OPEP force fields with the REMD, Chebaro et al. reproduced the thermal stability of GB1p.⁷⁰ The folding of GB1p series has been used in the optimization of force fields.^{68,71} The specially optimized force fields by Chen et al.⁶⁸ and by Irback and Mohanty⁷¹ are even able to discriminate the stability difference among the GB1p and its mutants.

We chose the GB1m2 mutant for the optimization and examination of our force field⁷² due to its stability and fast kinetics. It should be noted that in the optimization we only tried to reproduce the total β -sheet content by changing backbone–backbone HB strengths. Other factors such as native topologies were not considered at all. Since the β -hairpin folds slower than the α -helix, we found that the REMD simulations of the GB1m2 converge at a slower rate than the REMD simulations of the AK17 and Fs (Figure S1 in the SI), which is basically consistent with previous computational studies.⁶⁸ We therefore carried out three REMD simulations of the GB1m2 starting from three different sets of denatured conformations, two lasting for 300 ns and one lasting for 600 ns. The last 100 ns of each simulation were analyzed.

The generated conformations are clustered according to backbone and C_β atoms excluding two terminal residues with a RMSD cutoff of 0.2 nm, following the clustering scheme proposed by Daura et al.⁹¹ Figure 4a shows the superposition between the GB1p PDB structures (1pgb) and the representative structures of the largest clusters of the three REMD simulations at 300 K. The RMSDs of the three representative structures with respect to the experimental structure are 0.09, 0.09, and 0.14 nm. The slightly larger RMSD for the latter one is due to the fraying ends. The populations of the largest clusters for the three REMD simulations are 43.0%, 45.6%, and 45.5%, indicating a convergence for the sampling of native structures.

Another way to estimate native conformations is to count native backbone HBs.^{51,68,71} The number of native HBs in a given conformation is denoted by N_{nHB} . Figure 5a shows the probability distribution of N_{nHB} at 300 K. Assuming that a folded conformation has $N_{nHB} \geq 3$, the probabilities of folded conformations in the three REMD simulations are 43.2%, 45.4%, and 45.4%, respectively, consistent with the estimation by RMSD. The fractions of folded structures were

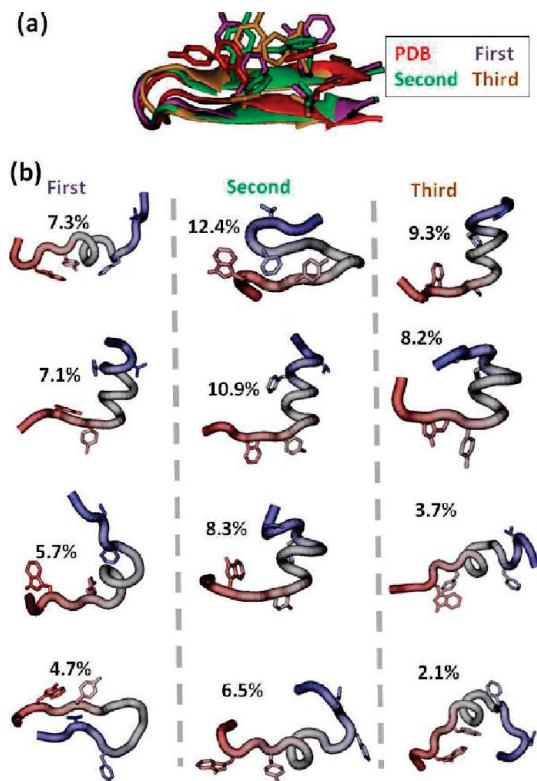


Figure 4. (a) Superimpositions of the native structure of the N-terminal hairpin of the protein G B1 domain (red) and the representative structures of the largest clusters of the three REMD simulations of the GB1m2 peptide at 300 K (purple, green, and brown). (b) The representative structures of the second to fifth most stable clusters of the three REMD simulations at 300 K.

computed for all replicas. The change in the folded fractions with temperature is shown in Figure 5b. Because the CG solvent developed by Marrink et al. has a frozen point of as high as ~ 290 K,⁷³ we could not obtain the melting curve at too low a temperature. However, since the folded fraction at 300 K is already close to one-half, we roughly estimated the T_m to be ~ 295 –300 K by interpolating the melting curve to 50% of folded fraction. Our estimated T_m is ~ 20 –25 K lower than experimental data (320 K).³³

There is an interesting puzzle about the hydrophobic interactions in the GB1p hairpin.⁹² There are four nonpolar side chains, Trp3, Tyr5, Phe12, and Val14. In both the crystal and NMR structures of the whole domain, all four residues contribute significantly to interstrand hydrophobic interactions to stabilize the hairpin.^{93,94} This can be seen clearly from our counting of the numbers of atomistic contacts between the side chains of the two strands (cross-strand atomistic contacts). As shown in Table 3, all four residues have at least 25 cross-strand atomistic contacts. Furthermore, the number of contacts between Trp3 and Val14 is 25, and the number of contacts between Tyr5 and Phe12 is 35. There are few contacts between Trp-Val14 and Tyr5-Phe12. Thus, Trp3-Val14 and Tyr5-Phe12 form two separated hydrophobic clusters. Nevertheless, a mutagenesis study revealed that the influence of mutation of Val14 on the thermal stability of the isolated GB1p hairpin is negligible compared with the others.⁹² Blanco et al. have also shown several interstrand

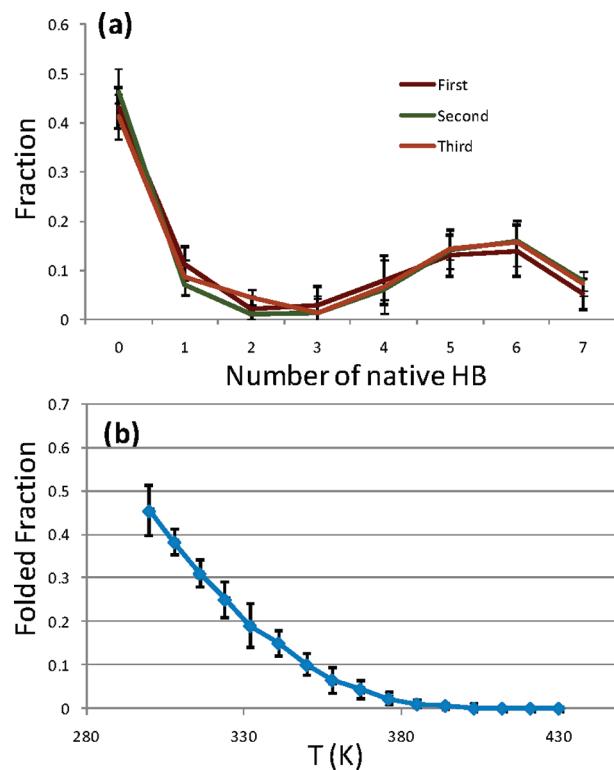


Figure 5. (a) Probability distribution of the number of native HBs in the three REMD simulations of the GB1m2 peptide. The native HBs are Pro7(O)–Gly10(NH), Asn6(O)–Lys11(NH), Asn6(NH)–Lys11(O), Thr4(O)–Thr13(NH), Thr4(NH)–Thr13(O), Glu2(O)–Thr15(NH), and Glu2(NH)–Thr15(O). (b) The melting curve of the REMD simulations of the GB1m2 peptide, which is the average of the three simulations. The error bars were estimated from block averages with a block size of 20 ns.

Table 3. Numbers of Cross-Strand Contacts Involved with Trp3, Tyr5, Phe12, and Val14 for the Folded and Unfolded Ensembles in the REMD Simulation of the GB1m2 Peptide at 300 K and the Native β -Hairpin from the Crystal Structure (PDB ID: 1pgb)

	Trp3	Tyr5	Phe12	Val14
folded ^a	22.9 ± 5.2	20.1 ± 4.1	34.5 ± 6.9	8.5 ± 2.7
unfolded ^a	5.8 ± 1.5	3.5 ± 1.0	8.8 ± 2.2	0.6 ± 0.2
PDB	31	35	41	25

^a The folded ensemble includes all of the conformations that have a RMSD of backbone and C_β atoms less than 0.2 nm; the other conformations are considered the unfolded structures. ^b For example, “22.9” here means that the folded ensemble has on average 22.9 atomistic contacts between the side chain of Trp3 and the side chains of hydrophobic amino acids on the opposing strand. A cutoff distance of 0.54 nm is used to count contacts.

NOEs in the NMR study of the GB1p hairpin, but no interstrand NOE between the Val14 side chain and other residues.³² In our folded ensemble (Table 3), the numbers of interstrand contacts involving Trp3, Tyr5, and Phe12 are 22.9 ± 5.2 , 20.1 ± 4.1 , and 34.5 ± 6.9 , respectively, which are still comparable to those in the crystal structures (31, 35, and 41, respectively). However the interstrand contact of Val14 is significantly reduced from 25 to 8.5 ± 2.7 , indicating that Val plays a less significant role in the cross-strand hydrophobic cluster. This agrees with experimental

observations. A visual investigation reveals that the loss of contacts of Val14 is because the side chain of Trp3 prefers to rotate to interact more with Tyr5 and Phe12, and consequently, the Trp3-Val14 cluster in the crystal structure is broken. Our results are also in line with a simulation study on this GB1p, in which only Val14 is detached but the other three aromatic side chains remain in the hydrophobic cluster at an early stage of an unfolding process.⁹⁵

Although the GB1p series are β -hairpin peptides, GB1p has been thought to possess a minor α -helical conformer. The simulation study of GB1p by Levy and co-workers⁴⁸ was able to estimate the free energy difference between the major β -hairpin state and the minor α -helix state at room temperature by applying the temperature-weighted histogram analysis method. According to their calculation, the α -helical state makes up ~8% of the population. The α -helical probabilities of our REMD simulations of GB1m2 are calculated (see Model and Methods) to be ~6.7–10.7%.^{48,96} The microstates accounting for this helical probability could be found from the most stable clusters in the REMD simulations. Figure 4b shows the second to fourth most stable clusters of the three REMD simulations. In all of the REMD simulations, there are persistent clusters with helical fragments spanning from Asn6 to Lys11 or even extending to the C-terminus. Using OPEP-REMD simulations, Chebaro et al. also found that GB1p has an α -helix spanning Asp6 to Thr11.⁷⁰ We suspect that the minor α -helical propensities of the GB1p series may be a consequence of their loop structures, since their six-residue loop requires the formation of an α -helical turn.

Folding of the Trpzip2 Peptide. Trpzip2 (SWTWENGK-WTKW), a 12-residue tryptophan zipper, is the smallest peptide adopting a unique β -fold.³⁵ With the high propensity of a type I' turn in the ENGK region and a characteristic structural motif of Trp-Trp cross-strand pairs, it has exceptional stability ($T_m \sim 345$ K) and fast folding kinetics ($\tau_f \sim 1.8 \mu\text{s}$).⁹⁷ The folding of Trpzip2 has been examined by numerous force fields.^{62,68,98,99} In the CHARMM-CMAP force field simulations with the GB solvent model, Chen et al. found that their REMD simulations are more difficult to converge for Trpzip2 than for the GB1p series despite the experimental fact that Trpzip2 folds much faster than the GB1p series.⁶⁸ In the study by Irback and Mohanty,⁷¹ their force field can fold the GB1p series correctly but cannot fold Trpzip2. This indicates that the folding of Trpzip2 is an interesting test of the ability of a force field to fold a β -sheet structure.

We performed two REMD simulations of the Trpzip2 peptide with 16 replicas from denatured conformations, one lasting for 120 ns and the other for 150 ns. It appears that convergence can be achieved after 50 ns of simulation (Figure S2 in SI), faster than in the simulations of the GB1m2 peptide (Figure S1 in SI). Thus, the last 70 ns of each REMD simulation of Trpzip2 were used for analysis. The sampled conformations were clustered according to the RMSD of backbone and C_β atoms of residues 2–11 with a cutoff of 0.2 nm. The representative structures of the largest clusters at 300 K in the two simulations are shown in Figure 6a, which are superimposed with the NMR structure of Trpzip2

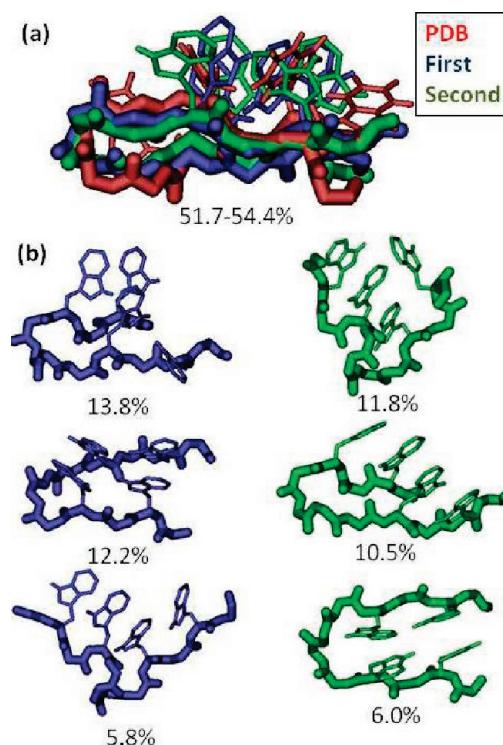


Figure 6. (a) Superposition of the native structure of Trpzip2 (red) and the representative structures of the largest clusters of the REMD simulations of this peptide at 300 K (blue and green, respectively). (b) The representative structures of the second to fourth most stable clusters of the two REMD simulations at 300 K.

(PDB ID: 1LE1).³⁵ Their RMSDs to the NMR structure are 0.14 and 0.16 nm. The populations of the largest cluster in the two simulations are 51.7% and 54.5%. Inspection of other top clusters (Figure 6b) of the two REMD simulations shows that the stabilities of the other minor non-native states are also consistent in the two simulations, suggesting that convergence has been achieved in the simulations.

The N_{nHB} analysis is also done for the REMD simulations of Trpzip2. The probability distributions of the two REMD simulations are shown in Figure 7a. Assuming the native structure with $N_{\text{nHB}} \geq 3$, the estimated populations of folded structures are 50.3% and 52.0% for the two REMD simulations, respectively, which are close to our estimation by clustering conformations. The melting curve of Trpzip2 is plotted by calculating the fraction of folded structures at different temperatures (Figure 7b), which gives a T_m of ~300–305 K. The calculated T_m is about 40–45 K lower than the experimental T_m for the Trpzip2 (345 K).³⁵

Our model underestimates the stability of the Trpzip2 native structure. In the folded ensemble of our simulations, the key stabilizing structural elements such as native HBs and type I' β -turns are well preserved. However, the packing of side chains of four Trp residues differs a lot from the NMR structure (Figure 6a). In the NMR structure, Trp11/Trp2 and Trp4/Trp9 form cross-strand pairs through an edge-to-face packing. To satisfy this packing motif, the hairpin is highly twisted.³⁵ In our folded structures, there is no edge-to-face packing among the Trp side chains. Instead, Trp11/Trp2 and Trp4/Trp9 form cross-strand pairs through ring

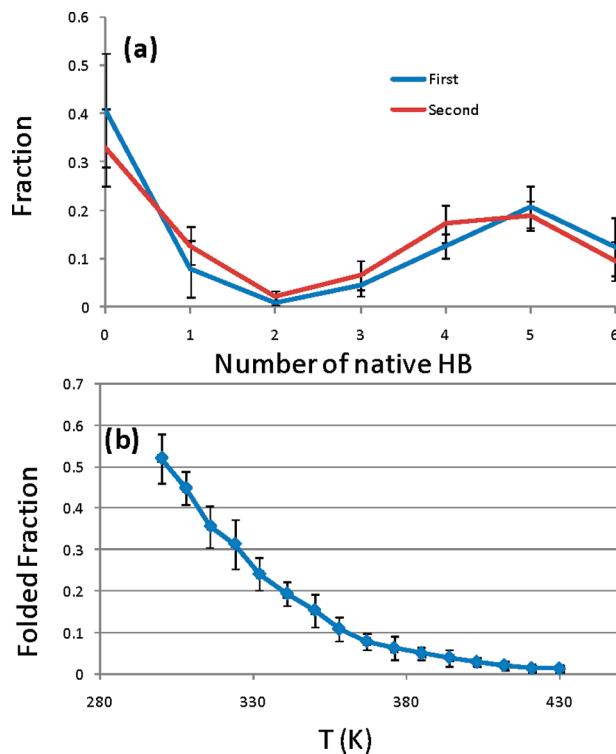


Figure 7. (a) Probability distribution of the number of native HBs in the two REMD simulations of the Trpzip2 peptide. The native HBs are Glu5(O)–Lys8(NH), Glu5(NH)–Lys8(O), Thr3(O)–Thr10(NH), Thr3(NH)–Thr10(O), Ser1(O)–Lys12(NH), and Ser1(NH)–Lys12(O). (b) The melting curve of the REMD simulations of the GB1m2 peptide, which is the average of the two simulations. The error bars were estimated from block averages with a block size of 20 ns.

stacking. As a result, the hairpin is less twisted. The loss of the experimental packing motif arises because the edge-to-face packing between aromatic rings is driven by electrostatic multipole interactions,¹⁰⁰ which is missing in our model. The aromatic groups in our model are only composed of van der Waal (vdW) particles. The free energy calculation in explicit solvent by Guvench and Brooks⁵⁰ revealed that for the Trpzip2 peptide, the edge-to-face packing between Trp side chains with the side chain multipole interactions provides ~ 4 kcal more stabilization than the stacked packing with only the side chain vdW interactions. This may account for the loss of stability in our Trpzip2 model. This may also be the reason why the force field by Irback and Mohanty,⁷¹ having a model of aromatic groups similar to ours, cannot fold the native β -hairpin topology of the Trpzip2 since it lacks the multipole interactions. Thus, inclusion of the multipole features of aromatic groups in explicit or implicit ways should be one of our next steps in the improvement of the force field.

Slow sampling for β -sheet peptides has been observed in previous computational studies. In a Monte Carlo (MC) study, Ulmschneider and Jorgensen⁶⁹ carried out MC simulations of two hairpin peptides, U(1–17)T9D¹⁰¹ and Trpzip2. They used a concerted-rotation MC move which is supposed to enhance sampling speed. For U(1–17)T9D, in eight simulations lasting for 20–80 million MC steps, only one simulation reached the native state. The other simulations

were trapped in non-native hairpins. For Trpzip2, despite much longer simulations (350–550 million steps), only two out of eight simulations reached the native state. Even with the REMD sampling method, Chen et al.⁶⁸ found that the native state of Trpzip2 was formed in one REMD simulation but was not formed in another. To examine the sampling speed for the β -sheet, we performed five 4.8–5 μ s normal MD simulations of Trpzip2 at 300 K starting from different denatured conformations. The folding of Trpzip2 is indeed slower than the folding of AK17. As shown in Figure 8a, four of the five simulations reached the native structures.

The folding time ranges from ~ 0.5 to $\sim 3.1 \mu$ s. The secondary structure analysis of the five trajectories reveals (Figure 8b) that before the native state is reached, the peptide experiences non-native β -hairpins and transient helical structures. In the fastest folding simulation (MD-1), no significant non-native secondary structures are observed. In the simulation (MD-5) that did not reach the native states, the system is trapped in several exchanging non-native hairpins. The non-native hairpins have fewer β -sheet structures than the native state, and their dwell time ranges from ~ 0.1 to about 1 μ s. However, once the native state is achieved, it lasts until the end of the simulation, indicating that the native state is much more stable than the non-native hairpins.

The four folding trajectories (MD 1–4) reveal two features of the folding pathways of Trpzip2. One is that all of the non-native secondary structures need to be completely unfolded before the folding of the native state starts, which is reflected by the secondary structure analysis (Figure 8b). The other feature is revealed by monitoring the structural change during the folding, which is shown in Figure 9. In all four simulations, the folding initiates at the type I' turn of Asn6-Gly7 and propagates toward the tails by forming cross-strand HBs sequentially. This folding mechanism is consistent with the “zip-out” model proposed by Munoz et al.¹⁰² in their experiments with the GB1p peptide, which was supported by later experimental and computational studies.^{103,104}

Simulations of the Trp-cage Peptide. The Trp-cage mini-protein (NLYIQ₅WLKD₁₀GPSSG₁₅RPPPS₂₀) is a designed peptide that behaves like large globular proteins (Figure 10a).³⁶ It contains an α -helix, a 3_{10} helix, and a polyproline II segment. Its tertiary structure is stabilized by a compact hydrophobic core centered at Trp6. The side chain of Trp6 is caged by the side chains of Tyr3, Leu7, Pro12, and Pro18. In addition, a distant HB between the side chain of Trp6 and the backbone carbonyl group of Arg16 and a salt bridge between the side chains of Asp9 and Arg16 are suggested to contribute to the stability of the Trp-cage. The temperature jump experiments show that the folding of the Trp-cage is extremely fast ($\tau_f \sim 4.1 \mu$ s),¹⁰⁵ which makes the Trp-cage an ideal model for computational study. Many simulations with explicit or implicit solvent have reproduced the folded structure of Trp-cage with excellent accuracy.^{9,52–58,68,70,71}

Because there are no attempts to fit the parameters of our force field by simulation of the Trp-cage, we think that the Trp-cage is a good case to test the transferability of our force field. We performed both REMD and normal MD simulations

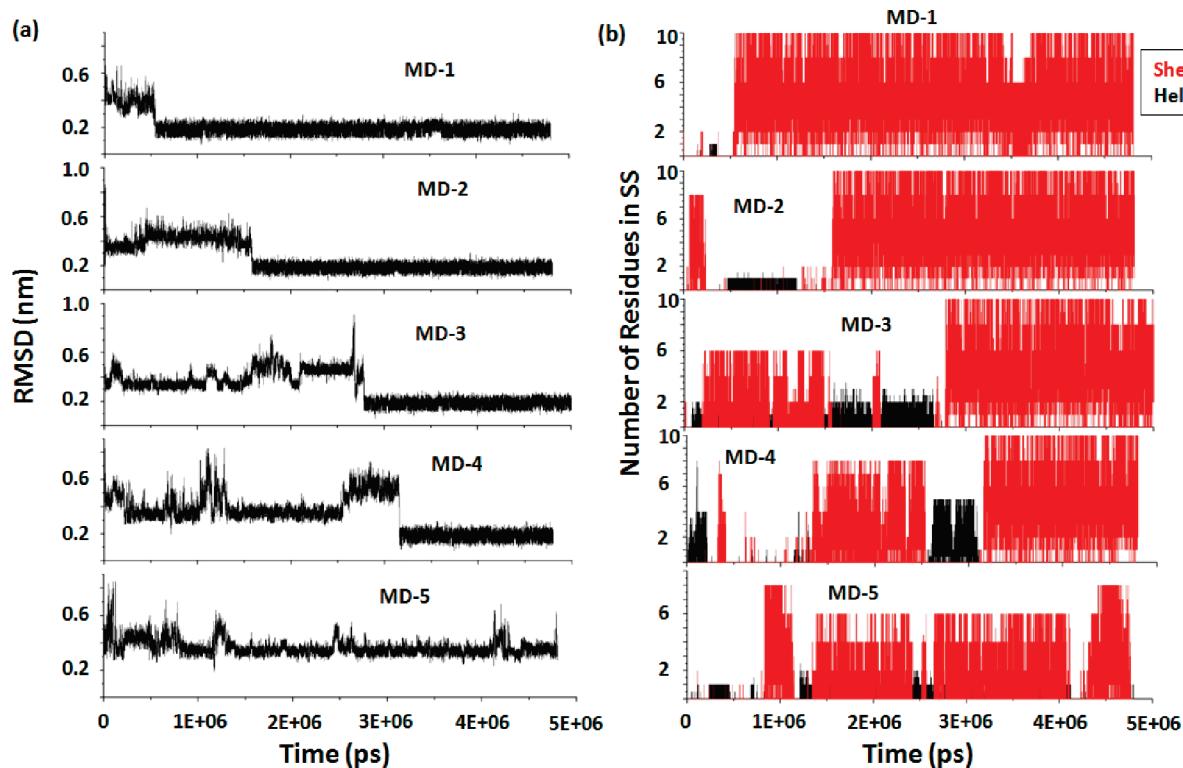


Figure 8. (a) RMSD with respect to the native structure (1LE1) of the normal MD simulations of Trpzip2. (b) The numbers of residues in helical (black) and sheet (red) structures.

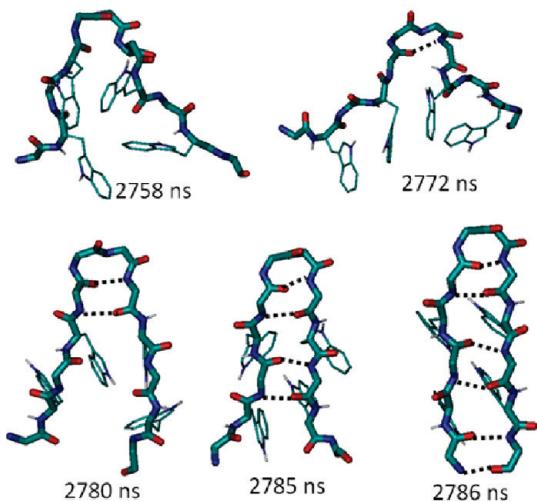


Figure 9. Representative pathway of the folding of Trpzip2 observed in our simulations.

for the Trp-cage peptide. The REMD simulation of the Trp-cage lasted for 180 ns, and the last 80 ns of simulations were used for structural analysis. The sampled conformations at 300 K were clustered according to the RMSD of backbone atoms of residues 2–19, discarding the fraying ends. The representative structures of the five most stable clusters are shown in Figure 10. The representative structure of the largest cluster has a RMSD of 0.20 nm with respect to the NMR native structure (PDB ID: 1L2Y).³⁶ In addition to a good fitting of the backbone structures, the native RMSD of all heavy atoms except for two terminal residues is ~ 0.27 nm for the representative structure of the largest cluster. In particular, the caged arrangement of the hydrophobic side chains is well reproduced. Assuming that the largest cluster

is folded, the fraction of folded conformations is 33.3%. We also applied the method of Garcia and co-workers to the calculation of a folded population, in which the native RMSD is calculated for each conformation and the conformation having a RMSD greater than a certain threshold is considered a folded state. We chose a threshold of 0.24 nm, slightly larger than Garcia et al.'s (0.22 nm).⁹ The folded population is estimated to be about 35.2%, consistent with our clustering analysis. Since the experimental probability of the folded structure is $\sim 70\%$ at 300 K,³⁶ our force field currently underestimates the stability of the native structure of Trp-cage.

We also carried out a 3 μ s normal MD simulation of Trp-cage at 300 K starting from a denatured conformation. The RMSD from the NMR structure is monitored during the simulations, which is plotted in Figure 11. Clearly, multiple folding and unfolding events occurred during the simulation. Discarding the first 500 ns of simulation, about 36.7% of sampled conformations have a RMSD from the NMR structure of less than 0.24 nm. This agrees very well with our REMD result. While the detailed folding mechanism of the normal MD trajectory is outside the scope of this paper, an estimation of the time scale of conformational change would be valuable. We estimated a folding time τ_f as the time the system takes to go deep into the folded basin ($\text{RMSD} < 0.14$ nm) after it enters the unfolded basin ($\text{RMSD} > 0.70$ nm). By this definition, six folding events are identified in our normal MD trajectory. The average folding time τ_f is approximately 270 ± 110 ns. Compared to the experimental τ_f of $4.1 \mu\text{s}$,¹⁰⁵ our model again has kinetics 1 order of magnitude faster than experimental data, similar to our observation in the normal MD of AK17.

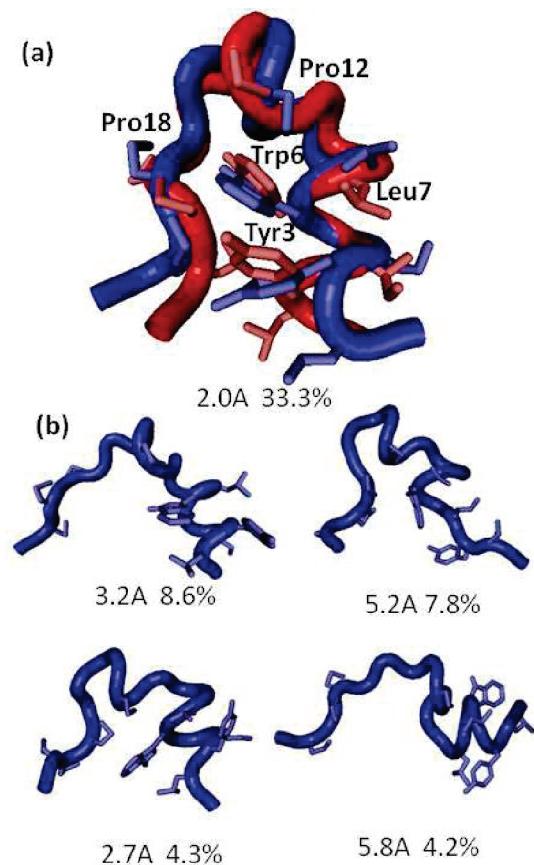


Figure 10. Representative structures, the probabilities, and the RMSD of the representative structures to the NMR structure (1L2Y) of the five most stable clusters (blue) of the REMD simulation of the Trp-cage at 300 K. (a) The representative structure of the most stable cluster is superimposed onto the NMR structure (red).

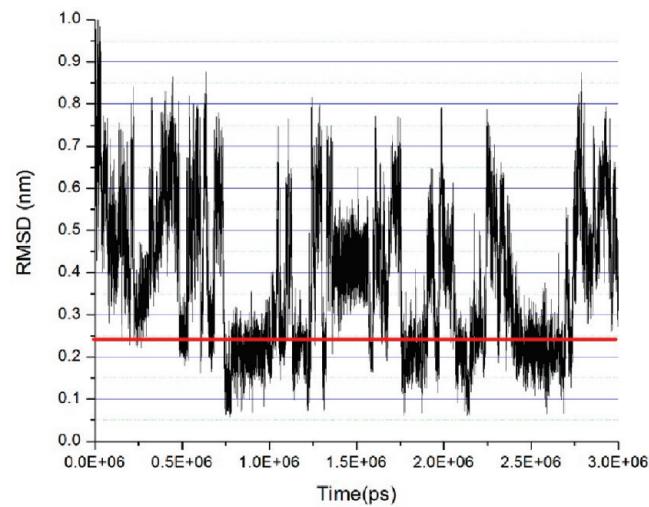


Figure 11. Change in the backbone RMSD (nm) to the NMR structure (1L2Y) during the normal MD simulation of the Trp-cage. The red line indicates the cutoff used for the calculation of the folded probability.

The HB ($\text{W6H}_\epsilon/\text{R16O}$) between the H_ϵ of the Trp6 side chain and the backbone carbonyl of Arg16 and the salt bridge (D9/R16) between the side chains of Asp9 and Arg16 have been widely studied.^{36,37} The $\text{W6H}_\epsilon/\text{R16O}$ interaction is well-supported by NOE signals.³⁷ The mutation of Trp6 by other

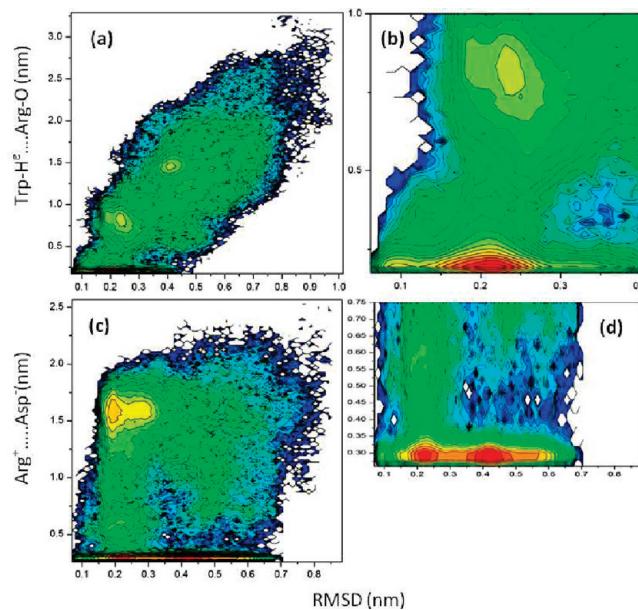


Figure 12. (a, b) The 2D statistical map of the backbone RMSD (nm) from the NMR structure against the $\text{W6H}_\epsilon/\text{R16O}$ distance (nm). (c, d) The 2D statistical map of the backbone RMSD (nm) from the NMR structure against the distance between the guanidine of Arg16 and the carboxylate of Asp (nm). Each contour line denotes 1.0 kJ/mol of free energy difference.

aromatic groups, including a naphthalene group, destabilizes the native structure by about 9–12.5 kJ/mol.¹⁰⁶ On the other hand, the D9/R16 salt bridge is less certain because the side chain/side chain NOEs across the bridging sites are absent.³⁶ The mutation that breaks the salt bridge only destabilizes the native structure by 3.4–6 kJ/mol.³⁷ In many MD simulations of the Trp-cage, the D9/R16 salt bridge is observed either as an interaction facilitating the folding^{52,107} or as a kinetic trap.^{56,108}

It is desirable to examine the roles of the $\text{W6H}_\epsilon/\text{R16O}$ HB and D9/R16 salt bridge in our folding simulations. To do so, we projected the normal MD trajectory onto a 2D map of the RMSD and the distance between W6H_ϵ and R16O (Figure 12a,b) and a 2D map of the RMSD and the distance between the guanidinium of Arg16 and the carboxylate of Asp (Figure 12c,d). It is clear that on the RMSD– $\text{W6H}_\epsilon/\text{R16O}$ map (Figure 12a) the $\text{W6H}_\epsilon/\text{R16O}$ distance correlates well with the RMSD. The $\text{W6H}_\epsilon/\text{R16O}$ HB can only be formed in the native basin (Figure 12b). On the contrary, there is no apparent correlation between the D9/R16 distance and the RMSD (Figure 12c). The salt bridge can form in both the native basin ($\text{RMSD} \sim 0.21$ nm) and the non-native basin ($\text{RMSD} \sim 0.42$ nm) (Figure 12d). Our observation is in line with the recent astonishing 100 μ s all-atom simulations with explicit solvent by Garcia and co-workers.⁹ They found that the probability of $\text{W6H}_\epsilon/\text{R16O}$ HB matches well with the folded fraction at all temperatures. However, at high temperatures, where the folded population is scarce, there is still 20%–30% D9/R16 salt bridge. A simple explanation for our observation is that since HBs have directionality, to form the $\text{W6H}_\epsilon/\text{R16O}$ HB, the side chain of Trp6 and the backbone of Arg16 should be arranged in a certain orientation, which can only be satisfied in the native conformation.

However, salt bridge interactions have little directionality, and the long side chain of Arg16 is quite flexible. Therefore, the D9/R16 salt bridge can be accommodated in various conformations and does not exclusively stabilize the native structure.

Conclusions

It is important to perform converged sampling of peptide folding to evaluate the quality of a force field.^{62,68,70,71} In this paper, we report our effort to fold peptides with our recently developed PACE force field in tandem with a CG solvent model. The peptides include AK, Fs, GB1m2, Trpzip2, and Trp-cage, which have been shown to fold in various topologies such as α -helix, β -sheet, and mixed helix/coil structures. Despite our force field not being optimized for the folding of these peptides, their native structures are all identified as the dominant conformations in our REMD simulations. Normal MD simulations have also been performed for AK17, Trpzip2, and Trp-cage. Convergence can be reached in microsecond simulations which give very similar results to the REMD simulations. It is noted that the simulated kinetics of AK17 and Trp-cage by the force field are about 1 order of magnitude faster than the experimental ones. We attribute this to the coarse-graining of the water solvent and the simplification of protein potentials. The force field appears to underestimate the stability of native structures of Trpzip2 and Trp-cage, indicating the need for further improvement of the force field, such as considering multipole interactions in aromatic rings. As pointed out in our preceding paper,⁷² our force field was parametrized to reproduce the results of all-atom force fields in a solvent-exposed environment. However, as our CG water model is simply LJ fluid, the dielectric screening effect of water may be implicitly incorporated into our parameters for interactions between polar sites. Thus, the current force field may only be suitable for the simulations of small peptides which have reasonably good exposure to a solvent. Our ongoing parametrization is underway to account for environment-dependent electrostatic interactions.

Acknowledgment. We are grateful to the Research Grants Council of Hong Kong (CA06/07.SC05, 663509) for financial support of this research.

Supporting Information Available: Figure S1, the change in the helical content and the folded fraction of the GB1m2 and AK17 peptides during the REMD simulations; Figure S2, the change in the helical content and folded fraction of the Trpzip2 peptide during the REMD simulations. This information is available free of charge via the Internet at <http://pubs.acs.org>

References

- Chen, J. H.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140.
- Van Gunsteren, W. F.; Dolene, J.; Mark, A. E. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149.
- Freiderichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. J. *J. Comput. Chem.* **2009**, *30*, 864.
- Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9.
- Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435.
- Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526.
- Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 4702.
- Day, R.; Paschk, D.; Garcia, A. E. *Proteins* **2010**, *78*, 1889.
- Khandogin, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16880.
- Liang, C.; Derreumaux, P.; Wei, G. *Biophys. J.* **2007**, *93*, 3353.
- Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, I. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
- Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
- Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
- Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10.
- Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.
- Fujitsuka, Y.; Chikenji, G.; Takada, S. *Proteins* **2006**, *62*, 381.
- Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394.
- Ding, F.; Tsao, D.; Nie, H. F.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010.
- Noid, W. G.; Chu, J. W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116.
- Han, W.; Wu, Y.-D. *J. Chem. Theory Comput.* **2007**, *3*, 2146.
- Han, W.; Wan, C.-K.; Wu, Y.-D. *J. Chem. Theory Comput.* **2008**, *4*, 1891.
- Brooks, B. R.; Bruccoleri, R.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
- Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
- Jorgenson, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.
- Marqsee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8898.
- Marqsee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.
- Lockhart, D. J.; Kim, P. S. *Science* **1993**, *260*, 198.
- Hughes, R. M.; Waters, M. L. *Curr. Opin. Struct. Biol.* **2006**, *16*, 514.
- Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584.

- (33) Fesinmeyer, R. M.; Hudson, F. M.; Anderson, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238.
- (34) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Anderson, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15483.
- (35) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578.
- (36) Neidigh, J. W.; Fesinmeyer, R. M.; Anderson, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425.
- (37) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Anerson, N. H. *Protein Eng. Des. Select.* **2008**, *21*, 171.
- (38) Struthers, M.; Ottesen, J. J.; Imperiali, B. *Folding Des.* **1998**, *3*, 95.
- (39) Struthers, M. D.; Cheng, R. C.; Imperiali, B. *Science* **1996**, *271*, 342.
- (40) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.
- (41) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Munoz, V. *J. Am. Chem. Soc.* **2008**, *130*, 7489.
- (42) Chowdhury, S.; Zhang, W.; Wu, C.; Xiong, G.; Duan, Y. *Biopolymers* **2003**, *68*, 63.
- (43) Zhang, W.; Lei, H.; Chowdhury, S.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 7479.
- (44) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *99*, 2782.
- (45) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472.
- (46) Zagrovic, B.; Jayachandran, G.; Millett, I. S.; Doniach, S.; Pande, V. S. *J. Mol. Biol.* **2005**, *353*, 232.
- (47) Bolhuis, P. G. *Biophys. J.* **2005**, *88*, 50.
- (48) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722.
- (49) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795.
- (50) Guvench, O.; Brooks, C. L. *J. Am. Chem. Soc.* **2005**, *127*, 4668.
- (51) Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R. M.; Wu, K.-P.; Baum, J. *J. Am. Chem. Soc.* **2007**, *129*, 4858.
- (52) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548.
- (53) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
- (54) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.
- (55) Schug, A.; Herges, T.; Wenzel, W. *Phys. Rev. Lett.* **2003**, *91*, 158102.
- (56) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280.
- (57) Zhan, L.; Chen, J. Z. Y.; Liu, W.-K. *Proteins* **2007**, *66*, 436.
- (58) Paschek, D.; Nymeyer, H.; Garcia, A. J. *Struct. Biol.* **2007**, *157*, 524.
- (59) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6465.
- (60) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *386*, 460.
- (61) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chem. Phys.* **2004**, *307*, 269.
- (62) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878.
- (63) Liwo, A.; Khalili, M.; Czaplski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260.
- (64) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008.
- (65) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000.
- (66) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310.
- (67) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362.
- (68) Chen, J.; Im, W.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2006**, *128*, 3728.
- (69) Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849.
- (70) Chebaro, Y.; Dong, X.; Laghaei, R.; Derreumaux, P.; Mousseau, N. *J. Phys. Chem. B* **2009**, *113*, 267.
- (71) Irback, A.; Mohanty, S. *Biophys. J.* **2005**, *88*, 1560.
- (72) Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y.-D. *J. Chem. Theory Comput.* **2010**, DOI: 10.1021/ct1003127.
- (73) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.
- (74) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Mariink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819.
- (75) Bond, P. J.; Holyyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 592.
- (76) Kasson, P. M.; Kelley, N. W.; Singh, N.; Vrljic, M.; Brunger, A. T.; Pande, V. S.; Biophys. J. *Proc Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11916.
- (77) Lockhart, D. J.; Kim, P. S. *Science* **1992**, *257*, 947.
- (78) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.
- (79) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.
- (80) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200.
- (81) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. *J. Am. Chem. Soc.* **2001**, *123*, 2388.
- (82) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669.
- (83) Miick, S. M.; Casteel, K. M.; Milhauser, G. L. *Biochemistry* **1993**, *32*, 8014.
- (84) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934.
- (85) Chakrabarty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843.
- (86) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786.
- (87) Williams, S.; Causgrove, T. P.; Gilman, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691.
- (88) Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2788.

- (89) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- (90) Lwin, T. Z.; Luo, R. *Protein Sci.* **2006**, *15*, 2642.
- (91) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236.
- (92) Kobayashi, N.; Honda, S.; Yoshii, H.; Munekata, E. *Biochemistry* **2000**, *39*, 6564.
- (93) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657.
- (94) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721.
- (95) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062.
- (96) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801.
- (97) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Acad. Natl. Sci. U.S.A.* **2004**, *101*, 4077.
- (98) Okur, a.; Strockbine, B.; Hornak, V.; Simmerling, C. *J. Comput. Chem.* **2003**, *24*, 21.
- (99) Chen, C.; Xiao, Y. *Bioinformatics* **2008**, *24*, 659.
- (100) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.
- (101) Zerella, R.; Chen, P. Y.; Evans, P. A.; Raine, A.; Williams, D. H. *Protein Sci.* **2000**, *9*, 2142.
- (102) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- (103) Du, D.; Zhu, Y.; Huang, C. Y.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915.
- (104) Zhang, J.; Qin, M.; Wang, W. *Proteins* **2006**, *62*, 672.
- (105) Qiu, L.; Pabit, S. A.; Roiberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952.
- (106) Barua, B.; Andersen, N. H. *Lett. Pept. Sci.* **2002**, *8*, 221.
- (107) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711.
- (108) Ding, F.; Buldyrev, s. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147.

CT100313A