Journal of Theoretical and Computational Chemistry
Vol. 13, No. 4 (2014) 1450026 (15 pages)

DOI: 10.1142/S0219633614500266

Folding simulation of Trp-cage utilizing a new AMBER compatible force field with coupled main chain torsions

Lirong Mou*, Xiangyu Jia†, Ya Gao†, Yongxiu Li†, John Z. H. Zhang†,‡ and Ye Mei†,‡,§

*Institute for Advanced Interdisciplinary Research
East China Normal University, Shanghai 200062, P. R. China

†Center for Laser and Computational Biophysics
State Key Laboratory of Precision Spectroscopy and Department of Physics
Institute of Theoretical and Computational Science
East China Normal University, Shanghai 200062, P. R. China

‡NYU-ECNU Center for Computational Chemistry at NYU Shanghai
Shanghai 200062, P. R. China

§ymei@phy.ecnu.edu.cn

Received 22 January 2014
Accepted 10 March 2014
Published 21 May 2014

A newly developed AMBER compatible force field with coupled backbone torsion potential terms (AMBER03^{2D}) is utilized in a folding simulation of a mini-protein Trp-cage. Through replica exchange and direct molecular dynamics (MD) simulations, a multi-step folding mechanism with a synergetic folding of the hydrophobic core (HPC) and the α-helix in the final stage is suggested. The native structure has the lowest free energy and the melting temperature predicted from the specific heat capacity C_v is only 12 K higher than the experimental measurement. This study, together with our previous study, shows that AMBER03^{2D} is an accurate force field that can be used for protein folding simulations.

Keywords: Trp-cage; AMBER force field; coupled backbone torsion; folding mechanism; melting temperature.

1. Introduction

Understanding the mechanism of protein folding is critical for fighting against protein misfolding related diseases and for protein engineering. Much progress in this field has been made in the past few decades.[1–3] As a complement to experiment, molecular dynamics (MD) is able to give an atomic description of the kinetic and thermodynamic properties of proteins covering a wide range of time scales, which are usually difficult to observe in experiment. Some peptides and small proteins have been studied[4–7] by MD simulations benefited from their small size and fast folding rate. Among them, a 20-residue mini-protein Trp-cage[8] (sequence: NLYIQ

§Corresponding author.

WLKDG GPSSG RPPPS) is an ideal system that has been well investigated by both theoretical[9–22] and experimental means.[23–29] Trp-cage consists of an α-helix from residue 2 to 8, a 3_{10}-helix from residue 11 to 14 and a polyproline II stretch. The indole ring of Trp6 is buried in the center of a hydrophobic core (HPC), which is formed by the side chains of Tyr3 and four proline residues (Pro12, Pro17, Pro18, and Pro19). The guanidine group of Arg16 and the carboxylic group of Asp9 can form a salt bridge. Some suggest that salt bridge may also contribute to the stability of protein structure.[16,30–32] While some experimental and theoretical studies showed that Trp-cage can maintain the folded state after the elimination of this salt bridge.[28,33] With continuous advancement of the force fields for protein, there is no difficulty in folding this short peptide. But to recapitulate some experimental observables, such as melting temperature, is still a challenge. More intriguing and more difficult questions are related to its folding mechanism. (1) What is the rate-limiting step in the process of folding? (2) Whether the folding of Trp-cage is a simple two state model or more complex? (3) How the environment confinement affects the folding landscape of this folding process?[34–36] (4) Whether the secondary-like structures form before the protein collapsed to a compact structure? Many studies have been devoted to find the answers. The NMR and CD experiments[8,27,37] suggested a simple two-state mechanism. While some other experiments suggested that the folding mechanism of the Trp-cage protein was even more complex and intermediates were involved before the occurrence of folding.[24,28,38] After 77 MD simulations, Duan *et al.* suggested that packing of the Trp6 side chain could be the rate-limiting step.[14,15] Tian *et al.*[34] noticed that the size, the polarization of the confinement and the effect of solvent competed with each other to determine the folding pathway. Zhou suggested that the α-helix was formed in the final stage.[39] On the contrary, some experimental results showed evidences of a helical structure in the denatured state of Trp-cage, suggesting an early formation of the α-helix.[24]

The controversy is resulted from the limitations of both the experiments and the computations. Theoretical prediction of protein folding pathway depends on the ergodicity of phase space and the reliability of force field employed. Due to the rapid development of computational techniques, the urgent requirement for an accurate force field is becoming more and more significant. Many studies have been devoted to the refinement of force fields. Best *et al.* optimized the additive AMBER[40] and CHARMM[41] all-atom force field. Li *et al.*[42] used to downhill simplex minimization algorithm to optimize the dihedral angel potential with respect to NMR measurements. This new force field, AMBER99SBnmr, is capable of giving improved results compared to those of AMBER99SB. Lindorff-Larsen *et al.*[43] optimized side-chain dihedral parameters of certain residues by matching with the high-level quantum-mechanical calculation. However, we deem that a major defect in current force fields consists in the simplistic functional form for main chain torsions. Due that the main chain torsions (φ and ψ) are not separable, employing coupled terms for the main chain torsions is essential, especially for delineating the potential energy barriers,

which are crucial to large conformational change. A new set of backbone torsion parameters utilizing coupled 2D main chain torsion terms has been developed in our group.[44] Preliminary examinations of its reliability for model peptides have shown that it is capable of generating more balanced secondary structure distribution than original AMBER force fields.

In this work, folding simulation of Trp-cage was carried out with this new AMBER compatible force field incorporating re-optimized 2D main chain torsion potential (AMBER03^{2D}). Other force field terms were directly extracted from AMBER03[45] force field. The simulated native state matched very well with the NMR structure and the predicted melting temperature was close to the experimental value. Meanwhile, a possible folding mechanism was proposed.

2. Method

2.1. *2D torsion*

In AMBER force field, the main chain torsions (φ and ψ, which are defined as the dihedrals formed by C–N–C_α–C and N–C_α–C–N, respectively) are treated separately. The potential energy of each torsion is expressed as a 1D Fourier expansion truncated after three or four terms. Due to the limited number of parameters, the parameterization of torsion term mainly focus on a small portion of space around the energy basis, which is deleterious for the study of large conformation change such as protein folding. Recently, a new functional form with 2D Fourier terms for main chain torsions has been proposed and parameterized by our group.[44] The alanine dipeptide (AD) was chosen as the model system for parameterization. The potential energy map for a series of (φ, ψ) pairs was calculated at M06 2X/aug-cc-pvtz//HF/6-31G** level using Gaussian 09 package.[46] By equalizing the QM and MM energies, the main chain torsion energy can be expressed as:

$$E_{\mathrm{mct}} = E_{\mathrm{int}} - E_{\mathrm{oth}} + G_{\mathrm{PCM}} - G_{\mathrm{GB}}, \tag{1}$$

where E_{int} and G_{PCM} are the internal energy and polar solvation energy calculated at QM level[47] employing SMD, a continuous solvation model.[48] It differs from our previous study in which the IEFPCM solvation model was employed. E_{oth} is the MM energy excluding the main chain torsions. G_{GB} is the generalized Born solvation energy.[49] E_{mct} can be written as:

$$E_{\mathrm{mct}}(\varphi, \psi) = \sum_{i=-N_\varphi}^{N_\varphi} \sum_{j=-N_\psi}^{N_\psi} C(m, n) e^{im\varphi} e^{in\psi}. \tag{2}$$

The expansion coefficient, the only parameters for the main chain torsion, can be obtained by imposing transformation

$$C(m, n) = \frac{1}{4\pi^2} \sum_{i=-N_\varphi}^{N_\varphi} \sum_{j=-N_\psi}^{N_\psi} E_{\mathrm{mct}}(\varphi, \psi) e^{-im\varphi} e^{-in\psi} \tag{3}$$

with $E_{\mathrm{mct}}(\varphi, \psi)$ calculated from Eq. (1). Other torsion terms around N–C_α and C_α–C bonds have their contributions also removed from the original AMBER force field and merged into 2D torsion terms. This new force field has been coded into the sander and pmemd modules of AMBER 11.

2.2. *Simulation protocol*

The simulation began with a linear structure of trp-cage. It was first optimized by 1000 steps of steep decent method followed by 500 steps of conjugate gradient method. The relaxed structure was gradually heated to 300 K in 100 ps. Then a 500 ps MD simulation at constant temperature was conducted to further relax the whole system. The final structure was chosen as the initial structure for all the 12 replicas in REMD simulations. Temperatures were set in a range from 261 to 542 K.[50] Generalized Born model[51] with an effective salt concentration of 0.2 M was utilized to mimic the solvation effect. Nonpolar solvation term was approximately represented by surface area term.[52] Integral time step was set to 1 fs. Temperature was regulated using Berendsen thermostat[53] with a coupling time constant of 1 ps. SHAKE algorithm[54] was used to constrain all the covalent bonds involving hydrogen atoms. Swaps were attempted every 0.25 ps and MD simulations were extended to 160 ns for each replica. Snapshots were saved every 0.25 ps. Besides REMD simulation, two 500 ns direct MD simulations at 300 K were also carried out to study the folding pathway.

2.3. *Analysis method*

Parallel Tempering Weighted Histogram Analysis Method (PTWHAM)[55] was used for the calculation of free energy. The potential energy density of states can be expressed as:

$$\Omega_m = \frac{\sum_{k=1}^{N_T} g_{mk}^{-1} H_{mk}}{\sum_{k=1}^{N_T} g_{mk}^{-1} N_k \Delta U \exp[f_k - \beta_k U_m]}, \tag{4}$$

where Ω_m is the density of state and g_{mk} is the statistical inefficiency which usually takes the value of 1. β_k is the Boltzmann constant and f_k is the free energy at the kth temperature. U_m is the energy and H_{mk} is the number of conformations with the potential energy values equal to U_m. N_k is all the conformations sampled at the kth temperature. N_T is the total number of replicas. Any observables at the temperature of interest can be calculated from the density of state. The root mean square deviation (RMSD) of the backbone atoms N, C_α and C and that of the HPC between simulated and the NMR structure (the first model in PDB entry 1L2Y) were computed. Other reaction coordinates (RCs), like radius of gyration (R_{g}), native contact (Q), heat capacity (C_{v}) and α-helix fraction were also analyzed. Native contact, a widely used metric, is defined by two non-neighboring residues separated by less than 7.0 Å.[56,57] There are 42 native contacts in the NMR structure. The formation of α-helix structure was determined by main chain dihedral angles ($-100° \leq \varphi \leq -30°$

and $-67^\circ \leq \psi \leq -7^\circ$). To avoid numerical discontinuity problem, the helicity is defined as:

$$\alpha_h = \frac{1}{\left\{1 + \left[\frac{1.5\times(\varphi+60^\circ)}{30}\right]^4\right\}\left\{1 + \left[\frac{1.5\times(\psi+45^\circ)}{25}\right]^4\right\}}. \quad (5)$$

When φ and ψ equal -60° and -45°, respectively, the helicity equals 1. When dihedrals are away from $(-60^\circ, -45^\circ)$, the helicity decays rapidly. All the ensemble statistical analysis was performed on the last 155 ns trajectory of each simulation.

3. Results and Discussion

During the REMD simulation, a wide range of temperatures were used. The efficiency for phase space sampling and the convergence speed were guaranteed by exchange ratio between neighboring temperatures. The potential energy distributions for different temperatures are shown in Fig. S1. There are adequate overlaps between neighboring temperatures and the exchange ratios are around 0.3. All the replicas have travelled in the temperature space back-and-forth for many times, and the folded states have been ever reached by most of the replicas, as shown in Figs. S2 and S3. When the folded structure is reached, the peptide stays in the vicinity of that structure and at low temperature for several to tens of nanoseconds, indicating the potential energy superiority of the folded structure under this interaction potential.

The distribution of the backbone RMSD from the first model of the NMR structures (Protein Data Bank entry 1L2Y) shows three peaks as depicted in Fig. 1(a). A configuration is considered to be in the unfolded state if its backbone RMSD is above 2.2 Å. This criterion is determined by the location of the distinct trough in the distribution of backbone RMSD and is consistent with that in the study by Day and coworkers.[19] There are two narrow peaks in the folded domain, which cover 63.3% of the configurations altogether. In the unfolded region, a broad distribution peaks at 3.25 Å and extends to over 8 Å. This peak possibly corresponds to an intermediate state with the peptide partially folded. Cluster analysis is an effective means to detect stable states. We used the *kclust* tool in MMTSB[58] to classify the conformations sampled at 300 K. The centroid structures of the top three clusters are shown in Fig. 1(b). The largest cluster holds 57% of the conformations. The backbone RMSD of the centroid structure is only 1.1 Å, which is definitely in the folded region. The peptide adopts a U shape and both the α-helix and the 3_{10}-helix are well folded. Superimposition of the simulated conformation with the smallest RMSD and the NMR structure is shown in Fig. 1(c). These two structures are well aligned for both the trace of the backbone atoms and the pack of the side chains. The TRP residue is well buried in a HPC. The second and the third clusters contain 18% and 15% of the conformations respectively. The centroid structures of these two clusters are in a near U shape but neither of them has α-helix completely formed. We also studied the positions of Trp-6 residue in the centroid structures of these three

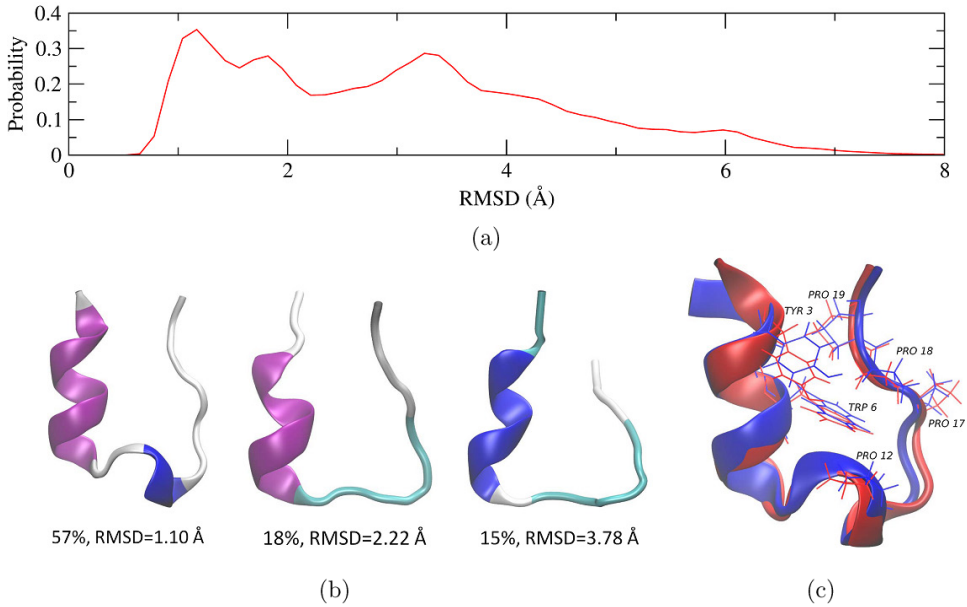

Fig. 1. (a) Probabilities of the RMSD distributions at 300 K; (b) Major conformation clusters of the trajectory at 300 K; (c) Overlay of the simulated conformation with the smallest RMSD at 300 K (red) and the NMR structure (blue).

main clusters. In the first cluster, the HPC has formed and Trp-6 is stable. But in the other clusters, the side chain of Trp-6 takes a wrong orientation and fluctuated dramatically. This suggested the nucleation-condensation mechanism and indicates one possible folding pathway: in the early stage, the tertiary and second structures are only partially folded; then the folding goes via the pack of Trp-6 in HPC followed by the full formation of α-helix.

In order to confirm this conjecture, we plot the free energy landscapes (FEL) mapped to several RCs. The FEL shown in Fig. 2 also supports the collapse-first mechanism. The HPC collapses in the early stage of the folding with the RMSD of the HPC decreases from 16 to 5 Å. Then the α-helix begins to grow. As we can see in Fig. 2, there are two barriers separating this FEL into three regions. In region III, The RMSD of HPC is near 4.0 Å and α-helix fraction distributes from 0.25 to 0.65. In region II, the RMSD of HPC translates from 4.0 Å to near 2.5 Å and the distribution of α-helix varies from 0.4 to 0.65. As we can see, from region III to II, further formation of α-helix does not occur and it is the optimization of HPC that drives this folding process. In region I, the RMSD of HPC is near 1.0 Å and α-helix fraction distributes from 0.5 to 0.8. So from region II to I, the growth of α-helix and further optimization of the HPC become synergetic. The α-helix fraction of the native structure is around 0.74, which is in the deepest free energy well.

We also carried out two direct MD simulations to study the folding process of Trp-cage. Each simulation ran for 500 ns at 300 K. For each trajectory, conformations

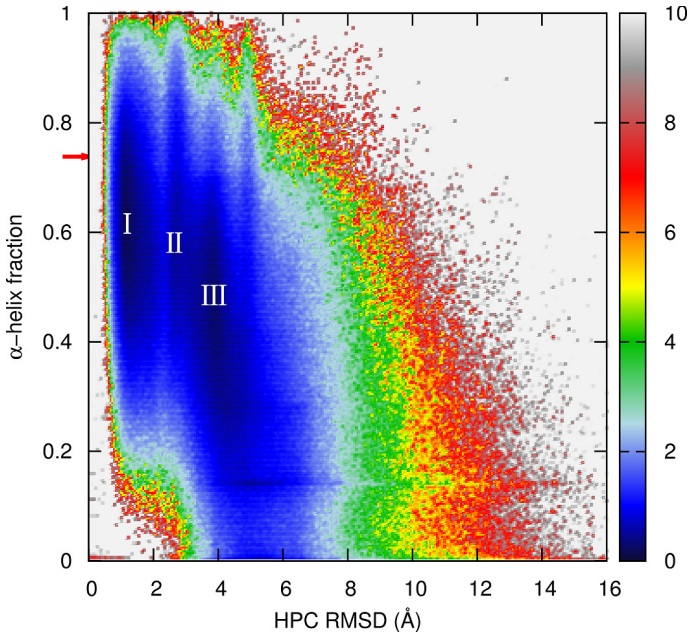

Fig. 2. Fitted free energy landscape in kcal/mol mapped to the RMSD from the NMR structure for the HPC and the fraction of α-helix at 300 K. The red arrow points to the location of the native state.

collected in the last 250 ns are used for analysis, in which the native state has been reached and folding/unfolding processes have occurred. In Fig. 3, the left panel shows the variation of RMSD over simulation time for the first trajectory. After cluster analysis, conformations belonging to the top five clusters are circled by different colors.

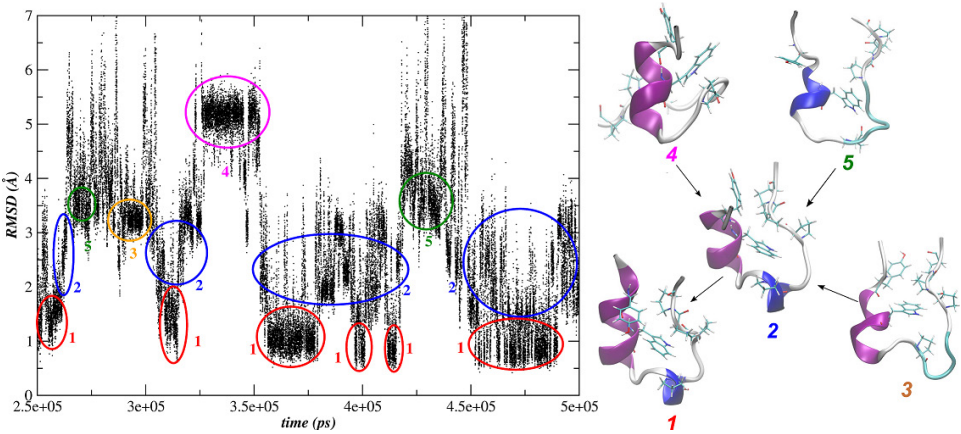

Fig. 3. The RMSD varies during the last 250 ns for the first direct MD trajectory and every colored circle encompasses the conformations belong to the same cluster (left part). The special structures nearest to centers of top five clusters and folding routes (black arrows) in this trajectory are displayed (right part).

Each cluster has a typical structure nearest to its center. These five typical structures are listed in the right side of Fig. 3. As we can see, for structures 3, 4 and 5 whose RMSD values are larger than 3 Å, their secondary structures are partially formed. But the indole ring of Trp-6 takes the wrong orientation. These structures may correspond to the conformations in region III of Fig. 2. To arrive at the native state, the structures in cluster 3, 4 and 5 must travel via cluster 2. RMSD of conformations in cluster 2 fluctuate largely from 1.5 Å to 3 Å. In this cluster, Trp-6 takes the correct orientation which further promotes the formation of HPC but the α-helix is still partially formed. This corresponds to region II in Fig. 2. HPC can stabilize the second structure and the second structure also contributes to the formation of HPC. This synergetic process drives protein going from cluster 2 to cluster 1 (the native state). For the second trajectory, results are described in Fig. 4. In this trajectory, structure 1 corresponds to the native state. Structure 2 has Trp-6 taken the right orientation and the α-helix partially formed. From unfolded state to the native state, structure 3 (RMSD >3.0 Å) must go through structure 2. In structure 5, all the second structures have almost completely formed but Trp-6 takes the wrong orientation. So, these secondary structure elements cannot be stabilized by HPC and Trp-cage must leave this well eventually to search for a more stable state.

In this study, a helix fragment is found in the unfolded state. This is consistent with some experiment measurements.[24,59] In some other computational studies, Xu *et al.*[17] used REMD with hybrid Hamiltonian to study the folding mechanism of Trp-cage and Chowdhury *et al.*[14] carried out 77 100-ns direct MD simulations for Trp-cage, they both found existences of helix structure in the unfolded state and the formation of HPC centered on Trp-6 was the rate-limiting step. Similar to their

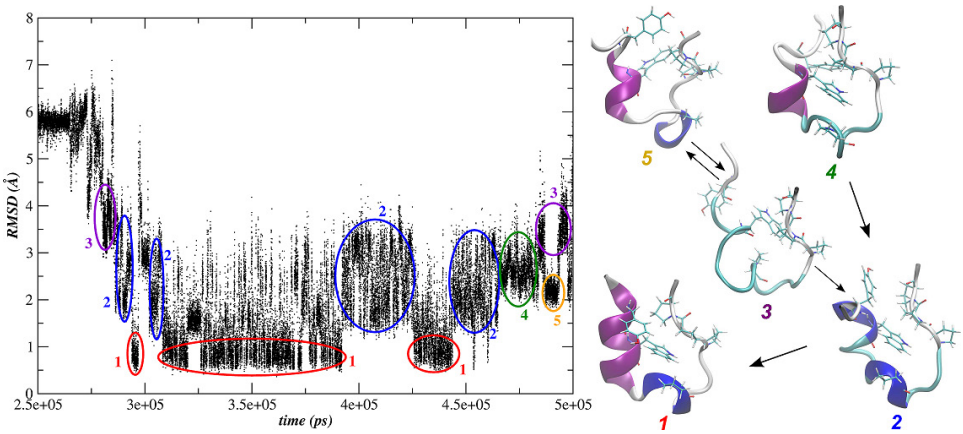

Fig. 4. The RMSD varies during the last 250 ns for the second direct MD trajectory and every colored circle encompasses the conformations belong to the same cluster (left part). The special structures nearest to centers of top five clusters and folding routes (black arrows) obtained from this trajectory are displayed (right part).

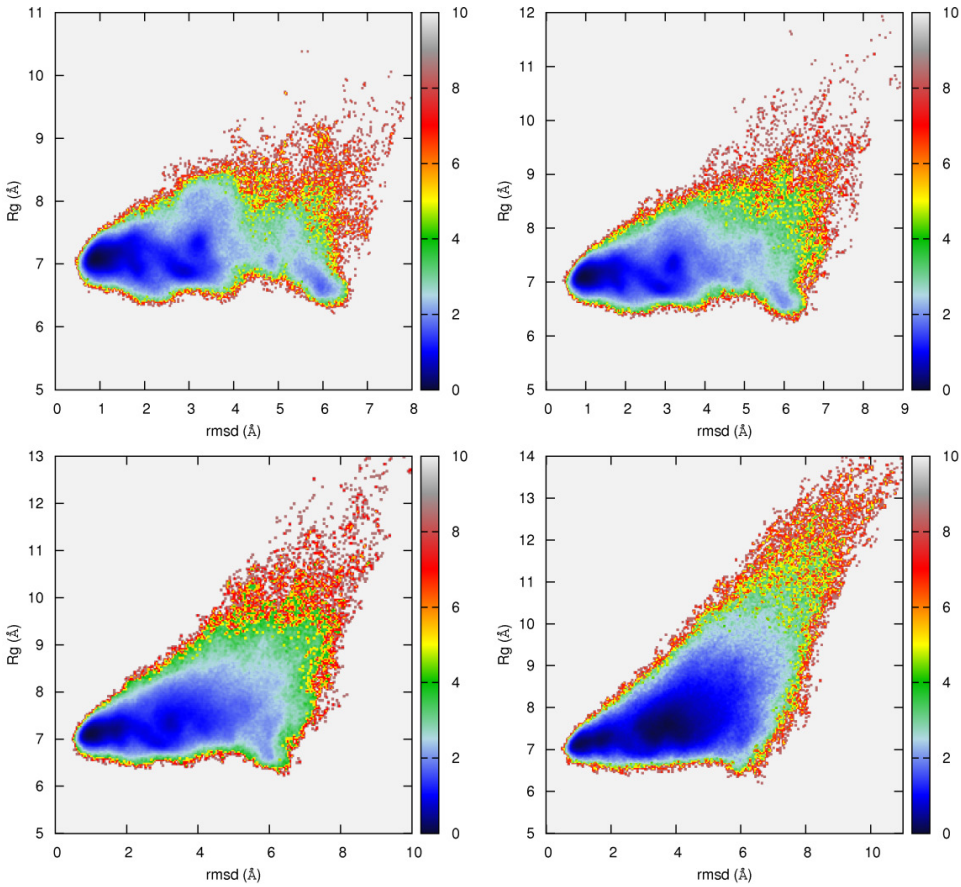

Fig. 5. Fitted free energy landscape in kcal/mol mapped to backbone RMSD and Rg at 261 K (upper left), 280 K (upper right), 300 K (low left) and 321 K (low right).

studies, we also suggest that correct orientation of Trp-6 and formation of HPC is an important step in folding from the intermediate state to native state.

The FEL of Trp-cage at four temperatures mapped to R_g and RMSD are shown in Fig. 5. The dominant free energy well at low temperature is located at the folded region with the backbone RMSD and R_g around 1.0 Å and 7.0 Å, respectively. A minor well is seen in the unfolded region with the backbone RMSD around 3 Å and a more compact structure than the native structure. Population shift from the folded region to the unfolded region can be detected with the increase of temperature, indicating a thermal unfolding event. The conformations are mainly distributed in the free energy well in the unfolded region at 321 K. Plot of the population in the folded region at all the temperatures is shown in Fig. 6(a) (see the black dots). The predicted melting temperature, at which the folded and unfolded conformations are equally populated, is 287 K, which is much lower than the experimental measurement (317 K).[26] However, the prediction may be biased if only a single RC is chosen

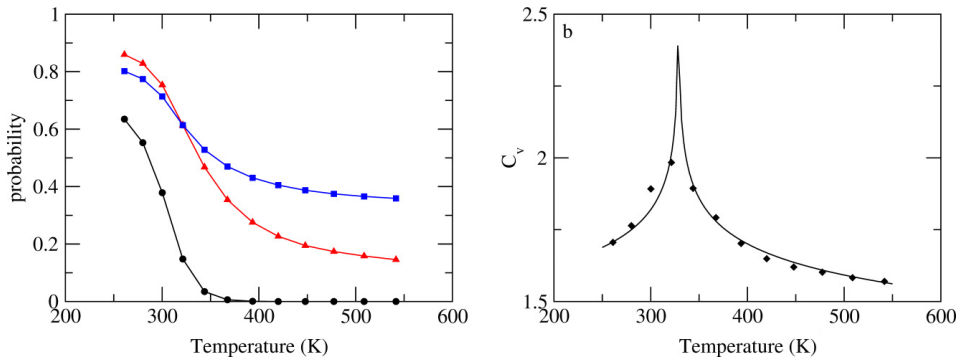

Fig. 6. (a) Variations of the population of folded state determined by (black dot) RMSD <2.2 Å, (red triangle) length of the native α-helix region and (blue square) the occurrence of the native contact in the temperature range; (b) The fitted specific heat curve.

to delineate the thermodynamics.[60] Therefore, the melting curves of the relative length of the native α-helix region and of the occurrence of the native contact are also studied. The melting temperature of the α-helix region is 354 K, which is much larger than that of the global unfolding. The native contact is even less labile, of which the melting temperature is another 22 K higher. Therefore, the predicted melting temperature shows strong dependence on the RC depicting the thermodynamics. Another way to obtain the melting temperature is via the temperature-dependence of specific heat capacity C_v, which can be calculated through the fluctuation of the total energy as

$$C_v = \frac{(\langle E^2\rangle - \langle E\rangle^2)}{RT^2}.$$

As the temperature approaches the melting temperature from both directions, there will be a clear spike in C_v theoretically. The fitted curve of C_v is shown in Fig. 6(b). Optimized parameters through Monte Carlo indicate a melting temperature of 329 K, which is only 12 K higher than the experimental measurement. Zhou *et al.*[39] used OPLS force field combined with explicit water model to study the melting temperature of Trp-cage by REMD simulation. The reaction coordinate they chose was native contact and the obtained melting temperature was higher than 400 K. In our study, melting temperature corresponding to native contact is about 339 K. Pitera *et al.*[9] used AMBER94 force field to study Trp-cage and the heat capacity was used to study folding-unfolding transition. They found a clear pike of heat capacity from 373 K to 433 K. Our calculation shows a better result by using heat capacity.

The FEL mapped to backbone RMSD and the distance between the titratable groups in Arg16 and Asp9 is shown in Fig. 7. It can be seen that the salt-bridge has an even distribution in both folded and unfolded state in the region with RMSD from 1 Å to 3 Å. Just as Zhou has pointed out,[39] this salt bridge can stable not only the

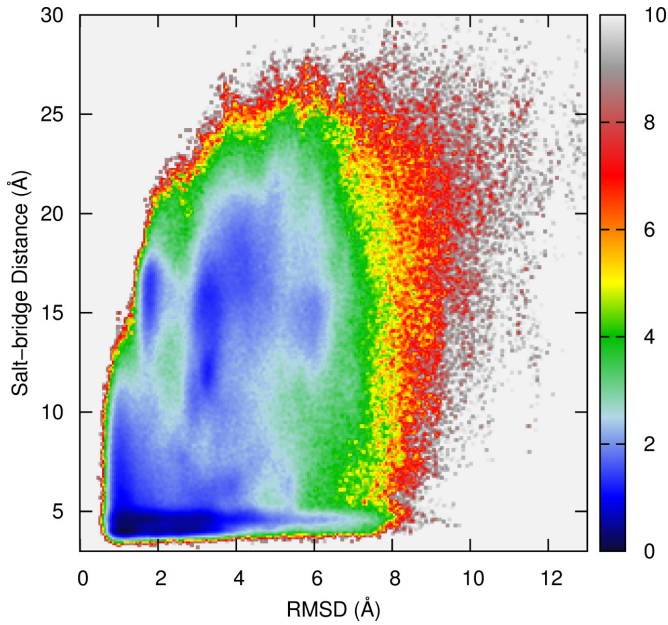

Fig. 7. Free energy landscape in kcal/mol mapped to backbone RMSD and the length of the salt bridge between the guanidine group of Arg16 and the carboxylic group of Asp9.

native structure but also the intermediate state. For the folding process from intermediate state to the native state, Trp-cage needs energy to break this salt-bridge.

4. Conclusions

Accurate description of protein folding from computational perspective mainly depends on the force field used. A critical limitation in current force fields is the secondary structure propensity. A new 2D torsion parameter is proposed in our group recently which implements cross terms of the main chain torsion dihedrals. When combined with AMBER force field, the issue of secondary structure biasing is alleviated. In this study, this new force field is utilized to study the folding of Trp-cage miniprotein through the replica exchange and direct MD simulations. This study suggests a multi-step folding mechanism with a synergetic folding of the HPC and the α-helix in the final stage. FEL and cluster analysis indicate the folding starts with a collapse of the structure followed by the formation of helices. Melting temperature determined by the divergence of the specific heat is 329 K, which is only 12 K higher than the experimental measurement.

In some previous studies,[61,62] it has been shown that GB model may overestimate charge–charge interaction and give a different free energy map about protein folding. In this study, GB7 model has been employed, which does a good job in describing salt bridge interaction. Besides, an effective salt concentration of 0.2 M is used to increase the screening effect. In our previous study,[44] we have noticed the model dependence

of conformation distribution on the specific solvation model. Normally, the force field should be tuned for a specific solvation model. The force field employed in this work has been tuned against GB models, therefore we use GB model in this folding simulation.

This work shows that the new force field is capable of folding this small peptide. Due to its well-balanced secondary structure propensity, it is capable of giving thermodynamic properties that are more in line with the experiment.

Supporting Information

The energy distributions for all the temperatures and the propagations of temperatures and RMSDs for all the replicas are shown in the supporting information.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 10974054, 20933002 and 21173082) and the Shanghai PuJiang Program (09PJ1404000). We thank Supercomputer Center of East China Normal University for CPU time support.

References

1. Best RB, Atomistic molecular simulations of protein folding, *Curr Opin Struct Biol* **22**:52–61, doi: 10.1016/j.sbi.2011.12.001 (2012).
2. Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA, The protein folding problem: When will it be solved? *Curr Opin Struct Biol* **17**:342–346, doi: 10.1016/j.sbi.2007.06.001 (2007).
3. Dill KA, MacCallum JL, The protein-folding problem, 50 years on, *Science* **338**:1042–1046, doi: 10.1126/science.1219021 (2012).
4. Zhang J, Qin M, Wang W, Folding mechanism of beta-hairpins studied by replica exchange molecular simulations, *Proteins: Struct Funct Bioinform* **62**:672–685, doi: 10.1002/prot.20813 (2006).
5. Shao Q, Gao YQ, Temperature dependence of hydrogen-bond stability in beta-hairpin structures, *J Chem Theory Comput* **6**:3750–3760, doi: 10.1021/ct100436r (2010).
6. Simmerling C, Strockbine B, Roitberg AE, All-atom structure prediction and folding simulations of a stable protein, *J Am Chem Soc* **124**:11258–11259, doi: 10.1021/ja0273851 (2002).
7. Shao Q, Shi J, Zhu W, Molecular dynamics simulation indicating cold denaturation of beta-hairpins, *J Chem Phys* **138**:085102, doi: 10.1063/1.4792299 (2013).
8. Neidigh JW, Fesinmeyer RM, Andersen NH, Designing a 20-residue protein, *Nat Struct Biol* **9**:425–430, doi: 10.1038/nsb798 (2002).
9. Pitera JW, Swope W, Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins, *Proc Natl Acad Sci USA* **100**:7587–7592, doi: 10.1073/pnas.1330954100 (2003).
10. Snow CD, Zagrovic B, Pande VS, The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations, *J Am Chem Soc* **124**:14548–14549, doi: 10.1021/ja028604l (2002).

11. Paschek D, Nymeyer H, Garcia AE, Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water, *J Struct Biol* **157**:524–533, doi: 10.1016/j.jsb.2006.10.031 (2007).
12. Mei Y *et al.*, Folding and thermodynamic studies of Trp-cage based on polarized force field, *Theor Chem Acc* **131**:1168, doi: 10.1007/s00214-012-1168-0 (2012).
13. Marinelli F, Pietrucci F, Laio A, Piana S, A kinetic model of Trp-cage folding from multiple biased molecular dynamics simulations, *PLoS Comput Biol* **5**:e1000452, doi: 10.1371/journal.pcbi.1000452 (2009).
14. Chowdhury S, Lee MC, Xiong GM, Duan Y, *Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution, *J Mol Biol* **327**:711–717, doi: 10.1016/s0022-2836(03)00177-3 (2003).
15. Chowdhury S, Lee MC, Duan Y, Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations, *J Phys Chem B* **108**:13855–13865, doi: 10.1021/jp0478920 (2004).
16. Hu Z, Tang Y, Wang H, Zhang X, Lei M, Dynamics and cooperativity of Trp-cage folding, *Arch Biochem Biophys* **475**:140–147, doi: 10.1016/j.abb.2008.04.024 (2008).
17. Xu W, Mu Y, *Ab initio* folding simulation of Trpcage by replica exchange with hybrid Hamiltonian, *Biophys Chem* **137**:116–125, doi: 10.1016/j.bpc.2008.08.002 (2008).
18. Zheng W, Gallicchio E, Deng N, Andrec M, Levy RM, Kinetic network study of the diversity and temperature dependence of Trp-cage folding pathways: Combining transition path theory with stochastic simulations, *J Phys Chem B* **115**:1512–1523, doi: 10.1021/jp1089596 (2011).
19. Day R, Paschek D, Garcia AE, Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein, *Proteins* **78**:1889–1899, doi: 10.1002/prot.22702 (2010).
20. Ding F, Buldyrev SV, Dokholyan NV, Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model, *Biophys J* **88**:147–155 (2005).
21. Juraszek J, Bolhuis PG, Sampling the multiple folding mechanisms of Trp-cage in explicit solvent, *Proc Natl Acad Sci USA* **103**:15859–15864, doi: 10.1073/pnas.0606692103 (2006).
22. Paschek D, Hempel S, Garcia AE, Computing the stability diagram of the Trp-cage miniprotein, *Proc Natl Acad Sci USA* **105**:17754–17759, doi: 10.1073/pnas.0804775105 (2008).
23. Halabis A, Zmudzinska W, Liwo A, Oldziej S, Conformational dynamics of the Trp-cage miniprotein at its folding temperature, *J Phys Chem B* **116**:6898–6907, doi: 10.1021/jp212630y (2012).
24. Ahmed Z, Beta IA, Mikhonin AV, Asher SA, UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein, *J Am Chem Soc* **127**:10943–10950, doi: 10.1021/ja050664e (2005).
25. Mok KH *et al.*, A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein, *Nature* **447**:106–109, doi: 10.1038/nature05728 (2007).
26. Streicher WW, Makhatadze GI, Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy, *Biochemistry* **46**:2876–2880, doi: 10.1021/bi602424x (2007).
27. Qiu LL, Pabit SA, Roitberg AE, Hagen SJ, Smaller and faster: The 20-residue Trp-cage protein folds in 4 mu s, *J Am Chem Soc* **124**:12952–12953, doi: 10.1021/ja0279141 (2002).
28. Neuweiler H, Doose S, Sauer M, A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate, *Proc Natl Acad Sci USA* **102**:16650–16655, doi: 10.1073/pnas.0507351102 (2005).

29. Culik RM, Serrano AL, Bunagan MR, Gai F, Achieving secondary structural resolution in kinetic measurements of protein folding: A case study of the folding mechanism of Trp-cage, *Angew Chem Int Ed* **50**:10884–10887, doi: 10.1002/anie.201104085 (2011).
30. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR, Strength and co-operativity of contributions of surface salt bridges to protein stability, *J Mol Biol* **216**:1031–1044, doi: 10.1016/s0022-2836(99)80018-7 (1990).
31. Waldburger CD, Schildbach JF, Sauer RT, Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol* **2**:122–128, doi: 10.1038/nsb0295-122 (1995).
32. 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* **2**:115–127, doi: 10.1021/ct050183l (2005).
33. Jimenez-Cruz CA, Makhatadze GI, Garcia AE, Protonation/deprotonation effects on the stability of the Trp-cage miniprotein, *Phys Chem Chem Phys* **13**:17056–17063, doi: 10.1039/c1cp21193e (2011).
34. Tian JH, Garcia AE, Simulation studies of protein folding/unfolding equilibrium under polar and nonpolar confinement, *J Am Chem Soc* **133**:15157–15164, doi: 10.1021/ja2054572 (2011).
35. Marino KA, Bolhuis PG, Confinement-induced states in the folding landscape of the Trp-cage miniprotein, *J Phys Chem B* **116**:11872–11880, doi: 10.1021/jp306727r (2012).
36. Samiotakis A, Cheung MS, Folding dynamics of Trp-cage in the presence of chemical interference and macromolecular crowding. I, *J Chem Phys* **135**:175101, doi: 10.1063/1.3656691 (2011).
37. Gellman SH, Woolfson DN, Mini-proteins Trp the light fantastic, *Nat Struct Mol Biol* **9**:408–410, doi: 10.1038/nsb0602-408 (2002).
38. Dyer RB, Ultrafast and downhill protein folding, *Curr Opin Struct Biol* **17**:38–47, doi: 10.1016/j.sbi.2007.01.001 (2007).
39. Zhou RH, Trp-cage: Folding free energy landscape in explicit water, *Proc Natl Acad Sci USA* **100**:13280–13285, doi: 10.1073/pnas.2233312100 (2003).
40. Best RB, Hummer G, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, *J Phys Chem B* **113**:9004–9015, doi: 10.1021/jp901540t (2009).
41. Best R *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles, *J Chem Theory Comput* **8**:3257–3273, doi: 10.1021/ct300400x (2012).
42. Li D-W, Brüschweiler R, Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins, *J Chem Theory Comput* **7**:1773–1782, doi: 10.1021/ct200094b (2011).
43. Lindorff-Larsen K *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins: Struct Funct Bioinform* **78**:1950–1958, doi: 10.1002/prot.22711 (2010).
44. Li YX *et al.*, A coupled two-dimensional main chain torsional potential for protein dynamics: Generation and implementation, *J Mol Model* **29**:3647–3657, doi: 10.1007/s00894-013-1879-8 (2013).
45. Duan Y *et al.*, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J Comput Chem* **24**:1999, doi: 10.1002/jcc.10349 (2003).
46. *Gaussian 09 B.01;* Gaussian, Inc.: Wallingford, CT, 2009.
47. Zhao Y, Truhlar D, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition

elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* **120**:215–241, doi: 10.1007/s00214-007-0310-x (2008).
48. Marenich AV, Cramer CJ, Truhlar DG, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, *J Phys Chem B* **113**:6378–6396, doi: 10.1021/jp810292n (2009).
49. Hawkins GD, Cramer CJ, Truhlar DG, Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium, *J Phys Chem* **100**:19824–19839, doi: 10.1021/jp961710n (1996).
50. Patriksson A, van der Spoel D, A temperature predictor for parallel tempering simulations, *Phys Chem Chem Phys* **10**:2073–2077, doi: 10.1039/b716554d (2008).
51. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A, Generalized born model with a simple, robust molecular volume correction, *J Chem Theory Comput* **3**:156–169, doi: 10.1021/ct600085e (2006).
52. Weiser J, Shenkin PS, Still WC, Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), *J Comput Chem* **20**:217–230, doi: 10.1002/(sici)1096-987x(19990130)20:2<217::aid-jcc4>3.0.co;2-a (1999).
53. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR, Molecular dynamics with coupling to an external bath, *J Chem Phys* **81**:3684–3690, doi: 10.1063/1.448118 (1984).
54. Ryckaert JP, Ciccotti G, Berendsen HJC, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Compu Phys* **23**:327–341, doi: 10.1016/0021-9991(77)90098-5 (1977).
55. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA, Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations, *J Chem Theory Comput* **3**:26–41, doi: 10.1021/ct0502864 (2007).
56. Juraszek J, Bolhuis PG, Rate constant and reaction coordinate of Trp-cage folding in explicit water, *Biophys J* **95**:4246–4257, doi: 10.1529/biophysj.108.136267 (2008).
57. Plaxco K, Simons K, Baker D, Contact order, transition state placement and the refolding rates of single domain proteins, *J Mol Biol* **277**:985–994, doi: 10.1006/jmbi.1998.1645 (1998).
58. Feig M, Karanicolas J, Brooks CL, MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graphics Model* **22**:377–395, doi: 10.1016/j.jmgm.2003.12.005 (2004).
59. Rovo P *et al.*, Structural insights into the Trp-cage folding intermediate formation, *Chem Eur J* **19**:2628–2640, doi: 10.1002/chem.201203764 (2013).
60. Krivov SV, Karplus M, Hidden complexity of free energy surfaces for peptide (protein) folding, *Proc Natl Acad Sci USA* **101**:14766–14770, doi: 10.1073/pnas.0406234101 (2004).
61. Li WF *et al.*, Effects of zinc binding on the conformational distribution of the amyloid-beta peptide based on molecular dynamics simulations, *J Phys Chem B* **111**:13814–13821, doi: 10.1021/jp076213t (2007).
62. Zhou RH, Berne BJ, Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proc Natl Acad Sci USA* **99**:12777–12782, doi: 10.1073/pnas.142430099 (2002).