

# RSC Advances



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This *Accepted Manuscript* will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# Important Roles of Hydrophobic Interactions in Folding and Charge Interactions in Misfolding of $\alpha$ -Helix Bundle Protein

Qiang Shao\*

Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai, 201203, China

\*To whom correspondence should be addressed. qshao@mail.shnc.ac.cn.

**Abstract** Integrated-tempering-sampling molecular dynamics simulation is utilized to investigate the folding of a 67-residue three- $\alpha$ -helix bundle,  $\alpha_3W$ . Reversible folding and unfolding can be observed in individual trajectories and a total of 28 folding events are achieved within 7  $\mu s$  simulation, giving sufficient sampling on the configuration space of protein folding. The native-like state having left-handed topology constitutes the largest fraction of the conformational ensemble sampled by the simulation. In addition, a misfolded state having mirror-image (right-handed) topology is observed with smaller population. The free energy landscape analysis demonstrates that the folding of  $\alpha_3W$  is initiated by the formation of  $\alpha$ -helical secondary structures and is followed by the assembling of folded  $\alpha$ -helices to construct tertiary structure. The “correct”  $\alpha$ -helix assembling which leads to native structure is mainly dominated by inter-helical hydrophobic interactions whereas the “incorrect” assembling which leads to misfolded mirror-image structure is highly affected by not only hydrophobic but also charge interactions. It is speculated on the basis of the present study on  $\alpha_3W$  and other studies on the B domain of protein A and  $\alpha_3D$  that the misfolding probability of  $\alpha$ -helix bundle protein is dependent on the strength of intra-protein hydrophobic and charge interactions: proteins containing stronger hydrophobic interactions but weaker charge interactions should have smaller misfolding probability. The importance of intra-protein hydrophobic interactions in preventing protein misfolding has been also seen in our previous studies on  $\beta$ -hairpins. Therefore, the present study along with our previous studies provide comprehensive, atomic-level picture of the folding/misfolding of  $\alpha$ -helix bundle and  $\beta$ -hairpin proteins.

**Keywords:** Folding and misfolding,  $\alpha$ -helix bundle, hydrophobic and charge interactions, molecular dynamics simulation, free energy landscape analysis

## 1. Introduction

The conformational specificity of a protein reflects its tendency to fold into a unique three-dimensional structure (or set of closely related structures). The failure to fold into functionally correlated native structure might result in severe physiological consequences. For instance, neurodegenerative disorders such as Parkinson's and Alzheimer's diseases are induced by the misfolding of protein to  $\beta$ -amyloid like structure which subsequently aggregates to form extracellular amyloid fibrils. Intriguingly, recent multiple molecular simulations have indicated that certain proteins/peptides can adopt alternative conformations in aqueous solution of which the most populated and thus experimentally detectable state is treated as native (folded) state whereas the less populated and experimentally undetectable one can be considered as misfolded state.<sup>1-9</sup> Unlike the irreversible protein misfolding and thus the incompatibility of folded and misfolded states of amyloid peptides, the folded and misfolded structures of these proteins with low conformational specificity can coexist in aqueous solution. For instance, multiple  $\beta$ -structured polypeptides (e.g., chignolin, the C-terminal  $\beta$ -hairpin fragment (residues 41-56) from the B1 domain of protein G (GB1p) and its mutants) have been observed to adopt a less populated "out-of-register"  $\beta$ -hairpin structure in addition to the experimentally measured native symmetric hairpin structure.<sup>1, 2, 4-6</sup> In comparison to the folded hairpin structure, the misfolded one has shorter turn region and oppositely orientated cross-strand hydrophobic side chains which are supposed to pack in same direction to form hydrophobic core cluster like in folded structure.

As popular models for *in vitro* investigation of protein folding, three- $\alpha$ -helix bundles are also geometrically interesting and has potential of misfolding. After the first two helices are aligned, the third one can be packed against either side of the plane involving the former two helices,

corresponding to left-handed or right-handed conformation.<sup>9</sup> The representative three- $\alpha$ -helix bundle, the B domain of protein A from *Staphylococcus aureus* (BdpA, 46 residues), has been indicated to adopt both native (left-handed) and topologically mirror-image (right-handed) conformations in several coarse-grained and all-atom simulations.<sup>7, 8, 10-13</sup> The molecular simulation with coarse-grained united-residue (UNRES) force field by Maisuradze *et al.* identified the mirror-image conformation of BdpA with very low population at room temperature.<sup>7</sup> More recent replica exchange molecular dynamics (REMD) and structure-based simulations by Noel *et al.* further quantified the mirror-image conformation with the occupation of 5% among folded structures of BdpA.<sup>8</sup> According to its small population, the mirror-image conformation of BdpA has been regarded as a misfolded state by the authors. In addition, both previous replica exchange Monte Carlo (REMC) simulation by Yang *et al.*<sup>14</sup> and more recent molecular dynamics (MD) simulation enhanced by dihedral-biased tempering method<sup>9</sup> indicated that another  $\alpha$ -helix bundle protein,  $\alpha_3W$  (67 residues) which has natively left-handed topology, could also misfold to non-native mirror-image conformation with observable population. The common observation of the mirror-image conformation in the simulations of multiple protein systems and implemented with various force fields suggests that it should be not the force field artifact.

Understanding the molecular mechanism underlying protein folding and (possible) misfolding is of fundamental importance to biology. The difficulty of experimental technologies in distinguishing the fold and misfolded states of abovementioned proteins and their respective transient states in the folding transition pathways, however, limits our approach to explore the folding/misfolding mechanism of these proteins. All-atom MD simulation can essentially provide detailed dynamic information of protein motion, but the low efficiency in sampling high-dimensional and rugged

configuration space of protein makes conventional MD simulation very computationally expensive.<sup>15</sup> Enhanced sampling methods are thus often used to improve the sampling efficiency of MD simulation, ensuring that the simulation can sufficiently sample protein configuration space.<sup>9, 16-22</sup> The associated free energy landscape analysis of simulation data can identify and energetically characterize protein folding/misfolding pathway.

Recently, we used integrated-tempering-sampling molecular dynamics (hereinafter referred to as ITS-MD) simulation<sup>23, 24</sup> to systematically investigate the folding and misfolding of  $\beta$ -hairpin systems.<sup>3, 5, 25-27</sup> The estimated folding kinetics and thermodynamics parameters (e.g., the folding time and individual backbone hydrogen bond stability of TRPZIP2, the cold denaturation tendency of five  $\beta$ -hairpins of MrH1, MrH4a, MrH3a, MrH3b, and MrH4b) are in well agreement with experimental data, indicating the validity of the simulation methodology in describing  $\beta$ -hairpin folding.<sup>3, 28-30</sup> More importantly, the misfolded states observed in most of  $\beta$ -hairpin systems are consistent with those reported by aforementioned simulation studies from other research groups.<sup>1, 2, 6</sup> The detailed free energy landscape analysis sheds light on the mechanism of  $\beta$ -hairpin misfolding which is induced by the alterable turn configuration and its competition with the cross-strand hydrophobic side chain interactions.<sup>25</sup> In the presence of strong  $\beta$ -turn-promoting sequence, the turn formation precedes over the formation of other two structural elements (the packing of cross-strand hydrophobic core cluster and backbone hydrogen bond assembling). On the other hand, the strong  $\beta$ -turn-promoting sequence might also have high risk for the turn adopting alternative conformations of which the native configuration does not but the non-native one does conflict with the packing of hydrophobic core cluster. As a result, the native turn configuration collaborates with the hydrophobic core to form correctly folded structure whereas the non-native turn configuration interrupts the

packing of the hydrophobic core to reach misfolded structure, especially for protein system containing weak hydrophobic interactions. In the presence of weak  $\beta$ -turn-promoting sequence and strong hydrophobic residues on the strand, the turn formation becomes the rate-limiting step and the folding process is initiated by the packing of hydrophobic core cluster, which strongly limits the approach to  $\beta$ -hairpin misfolding. For  $\alpha$ -helix bundle proteins, although previous studies suggested that right-handed conformation is dominated by charge interactions,<sup>9,31</sup> how charge interactions take into effect in the folding process to help the protein find a routine to target structure and what is the interplay between the charge interactions and the driving force for protein structure collapse are still unclear.

In this article, we utilized the same enhanced sampling simulation method (ITS-MD) to simulate the folding of a 67-residue  $\alpha$ -helix bundle protein,  $\alpha_3$ W (PDB code: 1LQ7<sup>32</sup>). In comparison to other  $\alpha$ -helix bundle analogues with relatively smaller sizes such as BdpA, the simulation studies on  $\alpha_3$ W are quite few, some of which have been performed using coarse-grained models.<sup>9,33,34</sup> The detailed folding mechanism of the protein is thus not clear yet. The present all-atom simulation captures not only low-energy native-like (left-handed) and misfolded mirror-image (right-handed) conformations but also various high-energy configurations along the folding transition pathway back and forth in individual trajectories. The captured mirror-image conformation has the same topology as that reported in previous simulations of  $\alpha_3$ W.<sup>9,14</sup> With sufficient protein folding/misfolding events collected in simulation, the folding and misfolding pathways of  $\alpha_3$ W can be identified through the free energy landscape analysis. It is observed that the folding pathway of  $\alpha_3$ W is mainly dominated by inter-helical hydrophobic interactions whereas the misfolding is also highly affected by inter-helical charge interactions. In addition, the present study suggests that the misfolding of

mirror-image conformation should be easier to occur for  $\alpha$ -helix bundle protein containing less hydrophobic but more charged residues. Such effects of intra-protein hydrophobic interaction strength on the misfolding tendency can be also seen for  $\beta$ -hairpin system.<sup>25</sup> Therefore, the present study together with our previous studies on  $\beta$ -hairpins<sup>3, 5, 25</sup> provide comprehensive, atomic-level picture of the folding and misfolding of  $\alpha$ -helix bundle and  $\beta$ -hairpin.

## 2. Computational Methods

All MD simulations were performed using AMBER11 package.<sup>35</sup> AMBER FF03 all-atom force field<sup>36</sup> was used to model protein and the recently improved generalized Born (GB) implicit solvent model, GB-Neck2 (*igb=8*),<sup>37</sup> was used to model solvent environment. The salt concentration was set to 0 and the default surface tension was 0.005 kcal/mol/Å<sup>2</sup>. The SHAKE algorithm<sup>38</sup> with a relative geometric tolerance of 10<sup>-5</sup> was used to constrain all chemical bonds. No non-bonded cutoff was used in any simulations.

Multiple independent simulation trajectories were performed. In each trajectory, the initial extended structure of the protein without residual secondary and tertiary structure elements was first subjected to 2500 steps of minimization. Then the system temperature was increased upon 300 K using velocity rearrangement from a Maxwell-Boltzmann distribution. After that, the system temperature was maintained at 300 K using the weak-coupling algorithm with a coupling time constant of 0.5 ps.<sup>39</sup> The long-time equilibrium simulation (production run) was performed with a step size of 0.002 ps and at the constant temperature of 300 K, but with a potential modified *via* the ITS method to encourage thorough sampling over a large potential energy range. The calculation data were collected every 2.0 ps and all data from the production run were used for data analysis.

The details of ITS method have been described previously.<sup>23, 24</sup> Briefly, a modified potential



energy is obtained from an integration function over temperature:

$$V' = -\frac{1}{\beta_0} \ln \sum_k f(\beta_k) e^{-\beta_k V} \quad (1)$$

where  $\beta_0 = \frac{1}{k_B T_0}$  ( $k_B$  is the Boltzmann constant and  $T_0$  is the temperature of system),  $V$  is the original potential function of the system under study.  $\beta_k$  is a series of temperatures that cover both low and high temperatures around the target one. The form of  $f(\beta)$  is taken as a sum of delta functions:

$$f(\beta) = \sum_{k=1}^N n_k \delta(\beta - \beta_k) \quad (2)$$

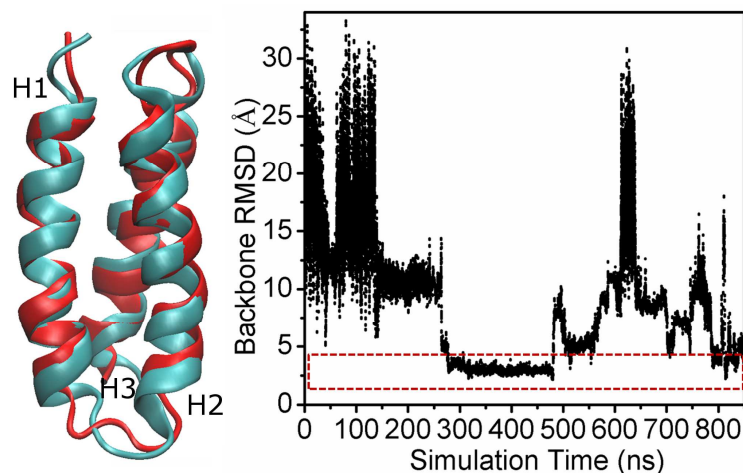
Then the integration in Eq. (1) could be written as a summation over a series of discrete temperature values ( $V' = -\frac{1}{\beta_0} \ln \sum_{k=1}^N n_k e^{-\beta_k V}$ ). Compared to conventional MD simulation, the sampled energy range in ITS-MD simulation could be greatly expanded depending upon the range of  $\beta_k$  used. In the present study, 300 temperatures ( $\beta_k$ ), evenly distributed in the range of 270-390 K, were used. A preliminary iteration process was employed to obtain converged values of discrete  $n_k$ 's, which were then used in the long-time production run to evenly sample the entire energy range. After the data were collected in production run using the modified potential with the set of  $f(\beta_k)$ 's, the thermodynamic properties of the simulated system at a desired temperature  $\beta$  were then calculated by reweighting of each term by timing a factor

$$e^{\beta(V'(r)-V(r))} = e^{-\beta V(r)} / p(V') = e^{-\beta V(r)} / \sum_{k=1}^N f(\beta_k) e^{-\beta_k V}.$$

### 3. Results

**3.1 One-dimensional Free Energy Profile Indicating the Presence of both Native and Mirror-image Conformations of  $\alpha_3W$ .** Twelve independent trajectories starting from diverse

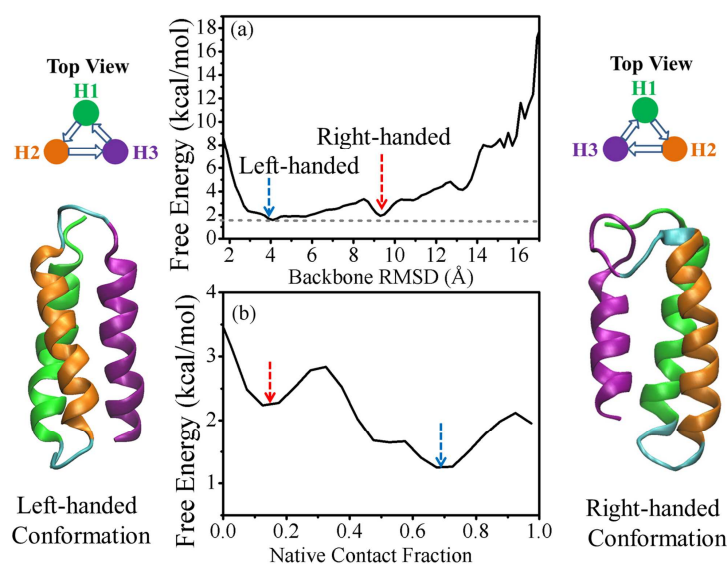
extended structures of  $\alpha_3W$  were run in order to obtain converged simulation results. Each trajectory lasted around 600 ns and the accumulated simulation time was around 7  $\mu s$ . Figure 1 shows time series of protein backbone heavy-atom root-mean-square deviation (RMSD) with respect to the NMR native structure of  $\alpha_3W$  in a typical trajectory. One can see that reversible folding (RMSD  $\leq 4.0$  Å) and unfolding events can be achieved in individual simulation trajectory. The sampled structures most fitting the NMR structure can reach the backbone heavy-atom RMSD of 1.78 Å (see left panel of Figure 1, the helices of H1-H3 are ordered from N- to C-terminal).



**Figure 1.** Left panel: Comparison of the simulated structure with smallest backbone heavy-atom RMSD values (blue) and the NMR structure (red) of  $\alpha_3W$ . Right panel: Time series of the backbone heavy-atom RMSD value in a typical trajectory of  $\alpha_3W$ .

The number of folding events is accumulated to  $\sim 28$  in all trajectories. The sufficient sampling of protein folding events allows us to statistically analyze the weight of various structures including the unfolded, transient, and folded structures along protein folding pathway. The one-dimensional free energy profile as a function of backbone heavy-atom RMSD was calculated at room temperature (Figure 2 (a)). As shown in this figure, the local free energy minimum at small RMSD position ( $\sim 3.90$  Å) has the lowest free energy, indicating that the native-like (left-handed) state of  $\alpha_3W$  is the

most populated. In addition, another local minimum with higher free energy is also presented at  $\text{RMSD} \approx 9.2 \text{ \AA}$ , corresponding to the mirror-image (right-handed) state. The free energy difference between the native and misfolded states is about  $-0.37 \text{ kcal/mol}$ , making the former state roughly 1.86 times more populated than the latter state. The free energy profile depicted as a function of another reaction coordinate, the fraction of native contacts of  $\alpha_3\text{W}$ , also indicates that the native state is more thermodynamically favorable than the misfolded state (the free energy difference is  $-0.88 \text{ kcal/mol}$ , Figure 2 (b)). In addition, a free energy barrier needs to be overcome for the transition from the misfolded state to native state ( $1.40 \text{ kcal/mol}$  in Figure 2 (a) and  $0.628 \text{ kcal/mol}$  in Figure 2 (b)).



**Figure 2.** One-dimensional free energy profiles as the function of (a) backbone heavy-atom RMSD value with respect to the NMR structure and (b) the fraction of native contacts of  $\alpha_3\text{W}$  at room temperature. The representative structures of native left-handed ( $\text{RMSD} \leq 4.0 \text{ \AA}$ ) and misfolded right-handed ( $\text{RMSD} \approx 9.2 \text{ \AA}$ ) conformations are shown at left and right sides, and their locations in the free energy profiles are pointed out by dash arrows, respectively.

To illuminate clearly the structure difference of the two states, we next calculated their contact

maps using the contact map analysis (CMA) server in SPACE tools.<sup>40</sup> One can see from Figure 3 that the main difference is in the residue-residue contacts between the helix H3 and the other two helices (see the regions circled in Figure 3). The observed mirror-image structure of  $\alpha_3W$  here has similar geometric features as that in previous explicit solvent MD simulation of the same protein:<sup>9</sup> both mirror-image structures have similar chirality (opposite to the native structure) and have less side chain-side chain contacts among hydrophobic residues than the native structure.

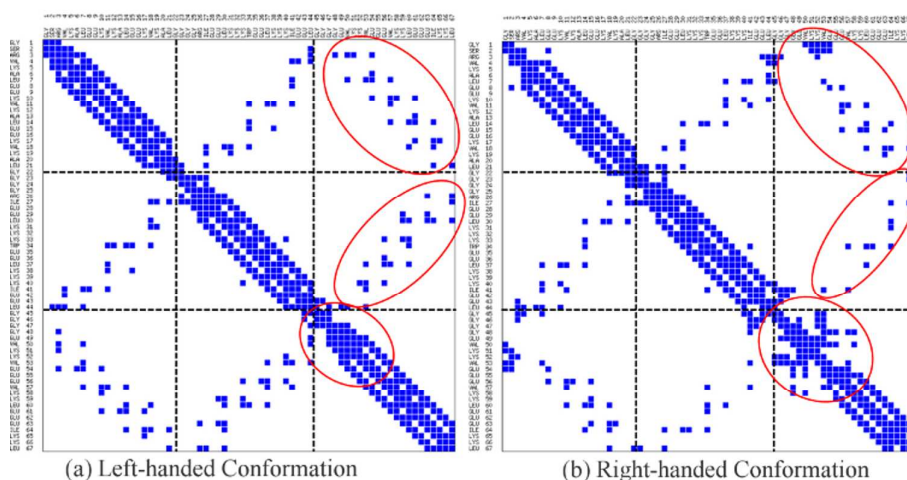


Figure 3. The contact maps of (a) left-handed and (b) right-handed structures of  $\alpha_3W$ . The contact map is calculated by the contact map analysis (CMA) server in SPACE tools.<sup>40</sup>

It is worth noting that the use of ITS-MD simulation can save the computational resource in comparison to other enhanced sampling methods. In our previous study on an Ala-Pro peptide,<sup>41</sup> we found that ITS-MD has an apparently higher efficiency than other bias potential and generalized ensemble methods, e.g., accelerated molecular dynamics (AMD) and replica-exchange molecular dynamics (REMD) simulations, in generating the thermodynamics parameters for the trans and cis interstate transition of the peptide. The present simulation ran 12 independent trajectories and each trajectory lasted ~600 ns. To achieve converged free energy profiles for a smaller  $\alpha$ -helix bundle protein, Engrailed Homeodomain protein (EnHD, 54 residues), Joshi and co-workers performed

implicit solvent REMD simulations of which either 30 parallel trajectories (replicas) covering the temperature range of 286-373 K<sup>42</sup> or 26 replicas covering the temperature range of 292-367 K<sup>43</sup> were implemented and the simulation per replica was lasted for 600 ns~2.0  $\mu$ s. The required computational resource and simulation time are greater than those in the present study. In addition, the coarse-grained canonical and multiplexed REMD simulation on the B domain of protein A from *Staphylococcus aureus* (BdpA, 46 residues) used 80 replicas to cover the temperature range of 250-500 K.<sup>7</sup> The use of explicit solvent REMD to study the folding of protein is even more computationally expensive, .e.g., a total of 64 replicas with the temperature range of 287-643 K were used by Noel et al. and 82 replicas with the temperature range of 277-548 K were used by Garcia et al. in their respective folding simulations on BdpA protein.<sup>8, 44</sup>

**3.2 Two-dimensional Free Energy Landscapes Indicating the Folding/Misfolding Mechanisms of Native and Mirror-image Conformations of  $\alpha_3$ W.** The folding of protein involves the formation of both secondary and tertiary structures. Since the native left-handed and misfolded right-handed conformations of  $\alpha_3$ W share same secondary structure elements ( $\alpha$ -helices of H1-H3 from N- to C-terminal), the molecular interactions which distinguish the two conformations are tertiary interactions (side chain-side chain interactions). The tertiary interactions can be divided into hydrophobic interactions from non-polar residues, favorite charge interactions from oppositely charged residues (salt bridges), and unfavorite charge interactions from residues carrying same types of charges. All of these interactions within individual subdomains (two subdomains formed by connecting helices, e.g., H1-H2, H2-H3, and one subdomain formed between two remote helices, H1-H3) are listed in Table 1 for both conformations. In comparison to the left-handed conformation, the right-handed conformation of  $\alpha_3$ W contains similar amount of favorite salt bridges but less

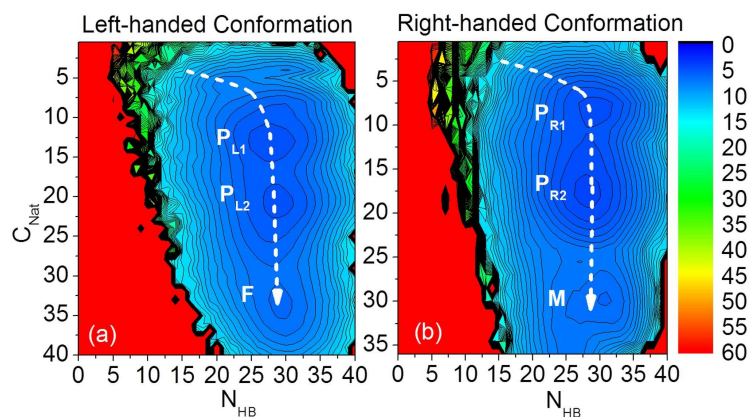
unfavorite charge interactions, indicating the intrinsic propensity of right-handed conformation.

Molecular Interactions	Native (Left-handed) Conformation			Misfolded (Right-handed) Conformation		
	H1-H2	H1-H3	H2-H3	H1-H2	H1-H3	H2-H3
Inter-Helical Hydrophobic Interactions	V4-I41	V4-V50	I27-I64	V4-L44	L7-V53	I27-I64
	V4-L44	L7-V50	I27-L67	A6-L44	L7-V57	I27-L67
	L7-L37	L7-V53	L30-L60	L7-L37	V11-V53	W34-V57
	L7-I41	L7-V57	L30-I64	L7-I41	V11-V57	W34-L60
	L7-L44	V11-V57	L30-L67	L7-L44	V11-L60	W34-I64
	V11-W34	L14-V57	W34-V57	L14-L30	L14-L60	L37-V57
	V11-L37	L14-L60	W34-L60	L14-W34	L14-I64	L37-L60
	V11-I41	L14-I64	L37-V53	L14-L37	V18-L60	I41-V53
	L14W34	V18-I64	L37-V57	V18-I27	V18-I64	I41-V57
	V18-I27	L21-I64	L37-L60	V18-L30	V18-L67	---
V18-L30	L21-L67	I41-V53	L21-I27	---	---	
V18-W34	---	L44-V53	---	---	---	
Inter-Helical Favorite Charge Interactions (Salt Bridges)	E15-K38	R3-E54	E29-K59	K17-E29	R3-E54	K38-E61
	---	K10-E61	K33-E56	---	K12-E56	---
	---	K17-E61	K40-E56	---	E15-K59	---
	---	---	---	---	K19-E63	---
Inter-Helical unfavorable Charge Interactions	K12-K38	R3-K51	E29-E63	K10-K33	E8-E56	---
	K19-K31	K10-K58	K33-K59	K17-R26	E15-E63	---
	---	---	K40-K52	---	---	---

**Table 1.** The inter-helical hydrophobic contacts, favorite and unfavorable charge interactions formed in individual subdomains in native left-handed and misfolded right-handed conformations of  $\alpha_3W$ .

To understand the interplay of the formation of secondary and tertiary structures in the folding/misfolding pathways of  $\alpha_3W$ , we used the total number of backbone hydrogen bonds formed ( $N_{HB}$ ) and the number of inter-helical side chain contacts ( $C_{Nat}$ ) to depict two-dimensional free energy landscapes for both native and mirror-image conformations. The former reaction coordinate is often used to describe secondary structure and the latter corresponds to tertiary structure. It is noteworthy that the detailed backbone hydrogen bonds are similar but the components of inter-helical side chain contacts are different for the two conformations.

As shown in Figure 4, three distinct local minima exist in these profiles. A large fraction of secondary structure contents are formed even in local minima containing fewest side chain contacts. Thus, the three local minima could be treated as two partially folded/misfolded states ( $P_{L1}$  and  $P_{L2}$  in Figure 4 (a) corresponding to the partially folded left-handed states and  $P_{R1}$  and  $P_{R2}$  in Figure 4 (b) corresponding to the partially misfolded right-handed states) and one folded (F) or misfolded (M) state. The connection of the three distinct states in each profile constructs the lowest free energy folding or misfolding pathway for protein. One can see that the number of backbone hydrogen bonds increases and reaches the maximum in very early stage of the folding of  $\alpha_3W$  ( $P_{L1}$  or  $P_{R1}$  state) where only a small portion of inter-helical side chain contacts are formed. The further folding of  $\alpha_3W$  involves the formation of more side chain contacts while the secondary structure is fully formed. Therefore, for either native or misfolded conformation, the formation of secondary structures precedes over that of tertiary structure, consistent with the framework mechanism: In the folding process of protein, its secondary structures are formed incipiently and independently of tertiary structure, then the formed secondary structure elements coalesce into native structure.<sup>45</sup> This observation is in well agreement with previous large-scale Monte Carlo simulation study of  $\alpha_3W$ .<sup>33</sup>



**Figure 4.** Two-dimensional free energy landscapes as the function of  $C_{\text{Nat}}$  and  $N_{\text{HB}}$  for (a) native left-handed and (b) misfolded right-handed conformations of  $\alpha_3\text{W}$  at room temperature. The contours are spaced at intervals of  $0.5 k_B T$ .

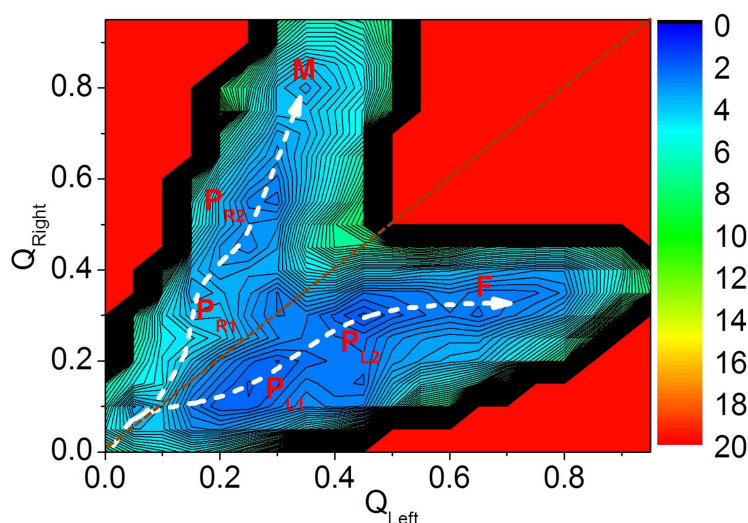
	Left-handed Conformation			Right-handed Conformation		
	$P_{L1}$ State	$P_{L2}$ State	F State	$P_{R1}$ State	$P_{R2}$ State	M State
Fraction of Hydrophobic Interactions	31.26%	43.50%	70.08%	32.07%	52.88%	79.16%
Fraction of Charge Interactions	18.50%	26.98%	35.66%	16.10%	65.10%	87.08%

**Table 2.** The average fraction of inter-helical hydrophobic and charge interactions formed in the three distinct states of the native and misfolded conformations of  $\alpha_3\text{W}$ , respectively.

The fractions of hydrophobic and charge interactions in each distinct state along the folding and misfolding pathways of  $\alpha_3\text{W}$ , respectively, were calculated. One can see from Table 2 that the fraction of hydrophobic interactions increases gradually in the transition pathways of both native and misfolded conformations. For left-handed conformation, the fraction of charge interactions is always smaller than that of hydrophobic interactions in all three states, suggesting that hydrophobic interactions but not charge interactions largely contribute for the folding of left-handed conformation.



In contrast, although only a small fraction of charge interactions are formed in the very early stage of the formation of right-handed conformation ( $P_{R1}$  state), most of charge interactions are formed in the middle and all of them are formed at the end of the misfolding pathway, strongly suggesting that charge interactions also play an important role in the misfolding of  $\alpha_3W$  to right-handed conformation. This observation is consistent with the speculation that charge interactions dominate the formation of right-handed conformation of  $\alpha$ -helix bundle.<sup>9, 31</sup>



**Figure 5.** Two-dimensional free energy landscapes as the function of the fraction of inter-helical side chain contacts for left- and right-handed conformations ( $Q_{Left}$  and  $Q_{Right}$ ) at room temperature. The contours are spaced at intervals of  $0.2 k_B T$ .

Next, to distinguish the folding pathway of left-handed conformation from the misfolding pathway of right-handed conformation, we calculated the free energy landscape as the function of the fraction of all inter-helical side chain contacts for both conformations ( $Q_{Left}$  and  $Q_{Right}$ ). The local minima corresponding to the partially folded/misfolded and fully folded/misfolded states are also presented in Figure 5. It is clear to be seen that the lowest free energy folding and misfolding pathways are roughly mirror symmetric along the diagonal axis. Therefore, the divergence of the folding of

left-handed and the misfolding of right-handed conformations of  $\alpha_3W$  occurs in the very early stage of  $\alpha_3W$  folding process. The average number of inter-helical hydrophobic contacts (the main content of tertiary interactions) formed within each subdomain in the distinct states along the folding/misfolding pathways of left-handed and right-handed conformations was calculated and organized in Table 3. For both conformations, more hydrophobic contacts are formed in H1-H2 subdomain than in the other two subdomains in the early stage of  $\alpha_3W$  folding ( $P_{L1}$  or  $P_{R1}$  state) and the further folding leads to the formation of more hydrophobic contacts in all subdomains.

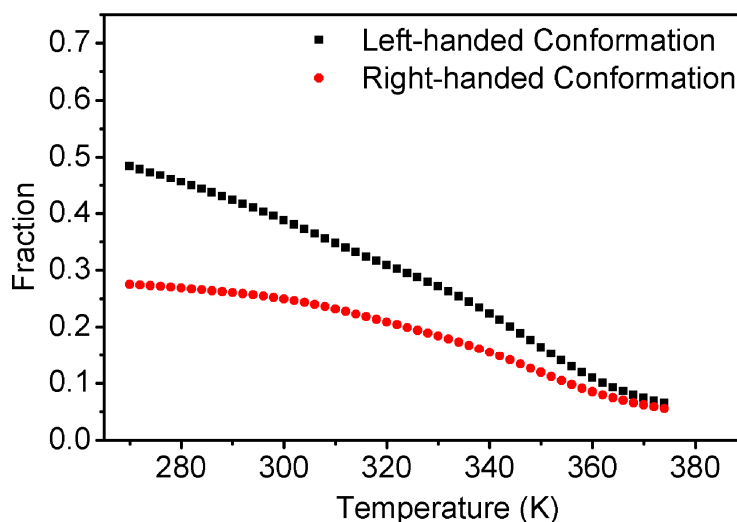
	Left Native Conformation			Right Non-native Conformation		
	$P_{L1}$ State	$P_{L2}$ State	$F_L$ State	$P_{R1}$ State	$P_{R2}$ State	$F_R$ State
$N_{HC, H1-H2}$	5.25	5.53	8.25	4.98	6.42	10.35
$N_{HC, H1-H3}$	2.84	6.50	8.52	2.88	3.02	5.84
$N_{HC, H2-H3}$	2.85	3.17	7.75	3.68	5.76	6.28

**Table 3.** The average number of inter-helical hydrophobic contacts formed in each subdomain for the three distinct states of the native and misfolded conformations of  $\alpha_3W$ , respectively.

### 3.3 Temperature Dependence of the Stability of Native and Mirror-image Conformations.

The molecular simulation with coarse-grained united-residue (UNRES) force field by Maisuradze *et al.* showed that the native and mirror-image conformations of BdpA protein display opposite temperature dependence in their stabilities: the population of the former conformation decreases whereas the population of the latter increases as the temperature is increased.<sup>7</sup> The REMD and structure-based simulations by Noel *et al.* demonstrated that the populations of both conformations of BdpA are decreased with temperature in general and the decrease in population of native conformation is much severer than that of mirror-image conformation.<sup>8</sup> Despite the discrepancy in

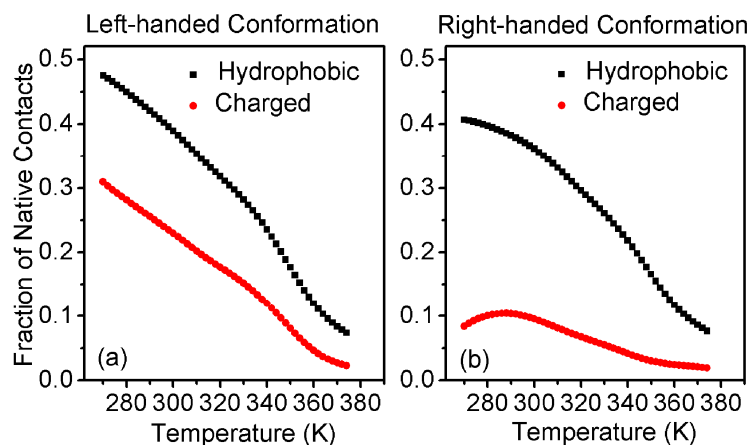
the description of temperature dependence of mirror-image conformation, both simulation studies suggested that the mirror-image conformation of BdpA has strong temperature resistance than native conformation. To evaluate the temperature resistance of the conformations with native and mirror-image topologies of  $\alpha_3W$ , we calculated the average fraction of inter-helical side chain contacts formed for both conformations at different temperatures (the number is averaged on the basis of all structures sampled in the simulation). One can see from Figure 6 that the contact fraction decreases with the temperature and the decreasing tendency of native conformation is more apparent than that of mirror-image conformation, suggesting that the former conformation has lower temperature resistance than the latter. As a result, the population difference between native and mirror-image conformations should be shortened as the temperature is increased.



**Figure 6.** The temperature dependence of the average fraction of inter-helical side chain interactions for left- and right-handed conformations of  $\alpha_3W$ .

To understand the different temperature resistance of the native and mirror-image conformations of  $\alpha_3W$ , we calculated the average fraction of inter-helical hydrophobic and charge interactions formed, respectively (Figure 7). Both kinds of interactions have their fraction decreased with

temperature in native conformation, corresponding to the low temperature resistance of this conformation. However, while the fraction of hydrophobic interactions keeps decreasing, the fraction of charge interactions in the non-native right-handed conformation keeps small but steady in a large temperature range. The small value of the fraction of charge interactions corresponds to the low population of non-native right-handed conformation and its steadiness explains the relatively strong temperature resistance of this conformation.



**Figure 7.** The temperature dependence of the average fraction of inter-helical hydrophobic and charge interactions for (a) left- and (b) right-handed conformations of  $\alpha_3W$ .

#### 4. Discussion and Conclusions

Recent molecular dynamics simulations suggested that protein misfolding could be a universal phenomenon among simple  $\beta$ -structured and  $\alpha$ -helical proteins/polypeptides.<sup>1-9</sup> The less population of misfolded structure in comparison to that of native structure, however, makes it difficult to be observed for experimental technologies, which leads to the difficulty in understanding the misfolding mechanism. Recently, we systematically investigated the folding and misfolding of  $\beta$ -hairpin using all-atom molecular dynamics simulation enhanced by integrated-tempering-sampling method, which gives reasonable explanation of  $\beta$ -hairpin misfolding.<sup>5, 25-28</sup> In the present study, to understand the

folding and unfolding mechanism of  $\alpha$ -helix bundle protein, we applied the same methodology on the folding simulation of a typical three- $\alpha$ -helix bundle,  $\alpha_3$ W.

In principle, the accuracy of theoretical simulation largely depends on the molecular force field used. The popular combination of AMBER FF03 force field and GB implicit solvent model, which has been widely used in many previous simulations on the members of  $\alpha$ -helix bundle proteins,<sup>42, 46-50</sup> was used here to model protein amino acids and water solvent, respectively. Although it is well-known that FF03 favors  $\alpha$ -helical conformations,<sup>51-54</sup> these abovementioned simulations can still provide reasonable description on the folding of  $\alpha$ -helix bundle proteins, with the calculated thermodynamics parameters (e.g., melting temperature and folding free energy) consistent with experimental data. Very recently, Simmerling and co-workers indicated that the molecular simulation combined with the improved generalized Born (GB) implicit solvent model, GB-Neck2 (*igb=8*),<sup>37</sup> can fold proteins with diverse topologies.<sup>55</sup> The combination of AMBER FF03 force field and GB-Neck2 model in the present simulation can not only fold  $\alpha_3$ W into its NMR experimentally identified native (left-handed) structure but also capture a misfolded mirror-image (right-handed) structure “hidden” behind the native one which has the similar topology as that observed in previous explicit solvent MD simulation<sup>9</sup>, indicating its efficiency in the sampling of protein configuration space. The former conformation occupies the largest fraction of the equilibrium conformational ensembles sampled by the simulation whereas the population of the latter is much smaller.

The detailed free energy landscape analysis indicates that the folding of  $\alpha_3$ W starts with the formation of all  $\alpha$ -helical secondary structures and is then followed by the assembling (packing) of folded  $\alpha$ -helices to construct the tertiary topology. The  $\alpha$ -helix assembling which involves the formation of inter-helical side chain contacts e.g., by hydrophobic and charge interactions, however,

could divaricate at the early stage of folding process, depending on which kinds of interaction makes dominant contribution in the structure assembling. The “correct” assembling which leads to native topology is solely dominated by inter-helical hydrophobic interactions whereas the “incorrect” assembling which leads to misfolded mirror-image structure is highly affected by not only inter-helical hydrophobic interactions but also charge interactions. The important role of hydrophobic interactions in folding native structure can be seen from the larger amount of hydrophobic interactions in native structure than those in mirror-image structure (Table 1) as well as the large fraction of hydrophobic interactions formed in the folding pathway (Table 2). Similarly, the large fraction of charge interactions in the misfolding pathway (Table 2) reflects its importance in  $\alpha_3W$  misfolding.

Considering the importance of hydrophobic interactions to folding and the importance of charge interactions to misfolding of  $\alpha_3W$ , one may speculate that the misfolding probability could be increased in the presence of less hydrophobic but more charged residues in amino acid sequence for  $\alpha$ -helix bundle protein. Using  $\alpha_3W$ , BdpA, and another  $\alpha$ -helix bundle protein  $\alpha_3D$  as examples, one can see that the hydrophobic and charged residues occupy 31.34% and 53.74% in the amino acid sequence of  $\alpha_3W$ , and the counterparts in BdpA occupy 41.30% and 26.08% and in  $\alpha_3D$  occupy 51.8% and 33.2%, respectively. Accordingly, the mirror-image structure of  $\alpha_3W$  has larger population than the mirror-image structure of BdpA (roughly 34.96% of folded structures of  $\alpha_3W$  are in mirror-image conformation for  $\alpha_3W$  as shown in the present study but the number is only 5% for BdpA in Ref. 8). In addition, both the explicit solvent MD simulation by Zhang and Ma<sup>9</sup> and our recent implicit solvent MD simulation<sup>50</sup> showed that no mirror-image structure is adopted for  $\alpha_3D$  which has the largest fraction of hydrophobic residues in the sequence among the three proteins

under study. Therefore, the misfolding probability of an  $\alpha$ -helix bundle protein could be evaluated by its amino acid sequence information. It is worth noting that our previous simulation suggested that the folding mechanism of an  $\alpha$ -helix bundle protein is also dependent on its sequence:  $\alpha$ -helical proteins containing higher percentage of hydrophobic residues than charged ones fold *via* nucleation-condensation mechanism (concurrent formation of secondary and tertiary structures) whereas proteins having opposite tendency in amino acid composition more likely fold *via* framework mechanism (incipient and independent formation of secondary structures before tertiary interactions).<sup>50</sup> In this scenario, both BdpA and  $\alpha_3$ D fold *via* nucleation-condensation mechanism whereas the folding of  $\alpha_3$ W follows framework mechanism.<sup>50</sup> Therefore, one may further speculate that the cooperative folding (nucleation-condensation) probably decrease the probability of misfolding. Combining the present study and our previous studies on  $\beta$ -hairpins,<sup>5,23-26</sup> we can find one common feature in the misfolding of  $\alpha$ -helix bundle and  $\beta$ -hairpin: proteins containing stronger intra-protein hydrophobic interactions have smaller probability to misfold.

**Acknowledgements** We thank the National Natural Science Foundation of China (Grant No. 21373258) and National Basic Research Program (Grant No. 2014CB910400) for financial support. The simulations were run at TianHe 1 supercomputer in Tianjin, and supercomputing facilities at The Texas Advanced Computing Center (TACC) at The University of Texas at Austin (Project ID: TG-MCB110130) and BlueBioU at Rice University.

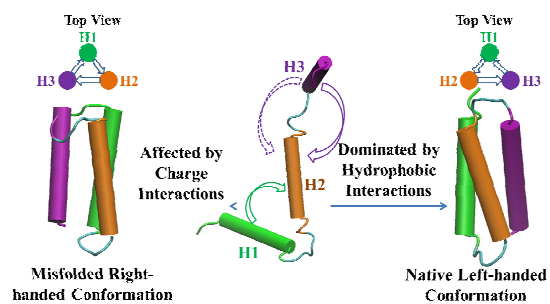
## References

1. R. B. Best and J. Mittal, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 11087-11092.
2. A. Bhattacharya, R. B. Best and J. Mittal, *Biophys. J.*, 2012, **103**, 596-600.
3. Q. Shao, J. Shi and W. Zhu, *J. Chem. Phys.*, 2013, **138**, 085102/1-085102/10.
4. Q. Shao, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2009, **130**, 195104/1-195104/6.
5. Q. Shao, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2011, **135**, 235104/1-235104/10.
6. P. Kuhrova, A. De Simone, M. Otyepka and R. B. Best, *Biophys. J.*, 2012, **102**, 1897-1906.
7. G. G. Maisuradze, A. Liwo, S. Oldziej and H. A. Scheraga, *J. Am. Chem. Soc.*, 2010, **132**, 9444-9452.
8. J. K. Noel, A. Schug, A. Verma, W. Wenzel, A. E. Garcia and J. N. Onuchic, *J. Phys. Chem. B*, 2012, **116**, 6880-6888.
9. C. Zhang and J. P. Ma, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 8139-8144.
10. J. Lee, A. Liwo and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 2025-2030.
11. A. Irback, F. Sjunnesson and S. Wallin, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 13614-13618.
12. K. Kachlishvili, G. G. Maisuradze, O. A. Martin, A. Liwo, J. A. Vila and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 8458-8463.
13. K. A. Olszewski, A. Kolinski and J. Skolnick, *Proteins*, 1996, **25**, 286-299.
14. J. S. Yang, W. W. Chen, J. Skolnick and E. I. Shakhnovich, *Structure*, 2007, **15**, 53-63.
15. J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror and D. E. Shaw, *Curr. Opin. Struct. Biol.*, 2009, **19**, 120-127.
16. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011-1021.
17. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1339-1350.
18. U. H. E. Hansmann, *Chem. Phys. Lett.*, 1997, **281**, 140-150.
19. G. Bussi, A. Laio and M. Parrinello, *Phys. Rev. Lett.*, 2006, **96**, 090601.
20. A. Mitsutake and Y. Okamoto, *Chem. Phys. Lett.*, 2000, **332**, 131-138.
21. F. G. Wang and D. P. Landau, *Phys. Rev. E*, 2001, **64**, 056101.
22. F. G. Wang and D. P. Landau, *Phys. Rev. Lett.*, 2001, **86**, 2050-2053.
23. Y. Q. Gao, *J. Chem. Phys.*, 2008, **128**, 064105/1-064105/6.
24. Q. Shao, W. L. Zhu and Y. Q. Gao, *J. Phys. Chem. B*, 2012, **116**, 13848-13856.
25. Q. Shao, J. Wang, J. Shi and W. Zhu, *J. Chem. Phys.*, 2013, **139**, 165103/1-165103/8.
26. Q. Shao and Y. Q. Gao, *J. Chem. Theory Comput.*, 2010, **6**, 3750-3760.
27. Q. Shao, H. Wei and Y. Q. Gao, *J. Mol. Biol.*, 2010, **402**, 595-609.
28. L. J. Yang, Q. Shao and Y. Q. Gao, *J. Phys. Chem. B*, 2009, **113**, 803-808.
29. A. J. Maynard, G. J. Sharman and M. S. Searle, *J. Am. Chem. Soc.*, 1998, **120**, 1996-2007.
30. R. B. Dyer, S. J. Maness, S. Franzen, R. M. Fesinmeyer, K. A. Olsen and N. H. Andersen, *Biochemistry*, 2005, **44**, 10406-10415.
31. J. W. Bryson, J. R. Desjarlais, T. M. Handel and W. F. DeGrado, *Protein Sci.*, 1998, **7**, 1404-1414.
32. Q. H. Dai, C. Tommos, E. J. Fuentes, M. R. A. Blomberg, P. L. Dutton and A. J. Wand, *J. Am. Chem. Soc.*, 2002, **124**, 10952-10953.
33. J. H. Meinke and U. H. E. Hansmann, *J. Comput. Chem.*, 2009, **30**, 1642-1648.



34. C. Czaplewski, S. Kalinowski, A. Liwo and H. A. Scheraga, *J. Chem. Theory Comput.*, 2009, **5**, 627-640.
35. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, AMBER11, University of California, San Francisco, 2010
36. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
37. H. Nguyen, D. R. Roe and C. Simmerling, *J. Chem. Theory Comput.*, 2013, **9**, 2020-2034.
38. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Computat. Phys.*, 1977, **23**, 327-341.
39. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
40. V. Sobolev, E. Eyal, S. Gerzon, V. Potapov, M. Babor, J. Prilusky and M. Edelman, *Nucleic Acids Res.*, 2005, **33**, W39-W43.
41. L. J. Yang, Q. Shao and Y. Q. Gao, *J. Chem. Phys.*, 2009, **130**, 124111/1124111/8.
42. S. Koulgi, U. Sonavane and R. Joshi, *J. Mol. Graphics Modell.*, 2010, **29**, 481-491.
43. V. Jani, U. B. Sonavane and R. Joshi, *J. Mol. Model.*, 2014, **20**, 2283.
44. A. E. Garcia and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 13898-13903.
45. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.*, 1990, **59**, 631-660.
46. H. X. Lei, C. Wu, Z. X. Wang, Y. Q. Zhou and Y. Duan, *J. Chem. Phys.*, 2008, **128**, 235105/1-235105/9.
47. H. X. Lei and Y. Duan, *J. Phys. Chem. B*, 2007, **111**, 5458-5463.
48. Q. Shao and Y. Q. Gao, *J. Chem. Phys.*, 2011, **135**, 135102/1-135102/13.
49. H. X. Lei, C. Wu, H. G. Liu and Y. Duan, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 4925-4930.
50. Q. Shao, *Journal of Physical Chemistry B*, 2014, **118**, 5891-5900.
51. S. Ono, N. Nakajima, J. Higo and H. Nakamura, *J. Comput. Chem.*, 2000, **21**, 748-762.
52. M. H. Zaman, M. Y. Shen, R. S. Berry, K. F. Freed and T. R. Sosnick, *J. Mol. Biol.*, 2003, **331**, 693-711.
53. J. Higo, N. Ito, M. Kuroda, S. Ono, N. Nakajima and H. Nakamura, *Protein Sci.*, 2001, **10**, 1160-1171.
54. N. Kamiya, J. Higo and H. Nakamura, *Protein Sci.*, 2002, **11**, 2297-2307.
55. H. Nguyen, J. Maier, H. Huang, V. Perrone and C. Simmerling, *J. Am. Chem. Soc.*, 2014, **136**, 13959-13962.

## Table of Contents Entry



An enhanced-sampling molecular dynamics simulation is presented to quantitatively demonstrate the important roles of hydrophobic and charge interactions in the folding and misfolding of  $\alpha$ -helix bundle protein, respectively.