

# SCIENTIFIC REPORTS



OPEN

## The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: a microscopic view

Received: 08 September 2015

Accepted: 14 December 2015

Published: 18 January 2016

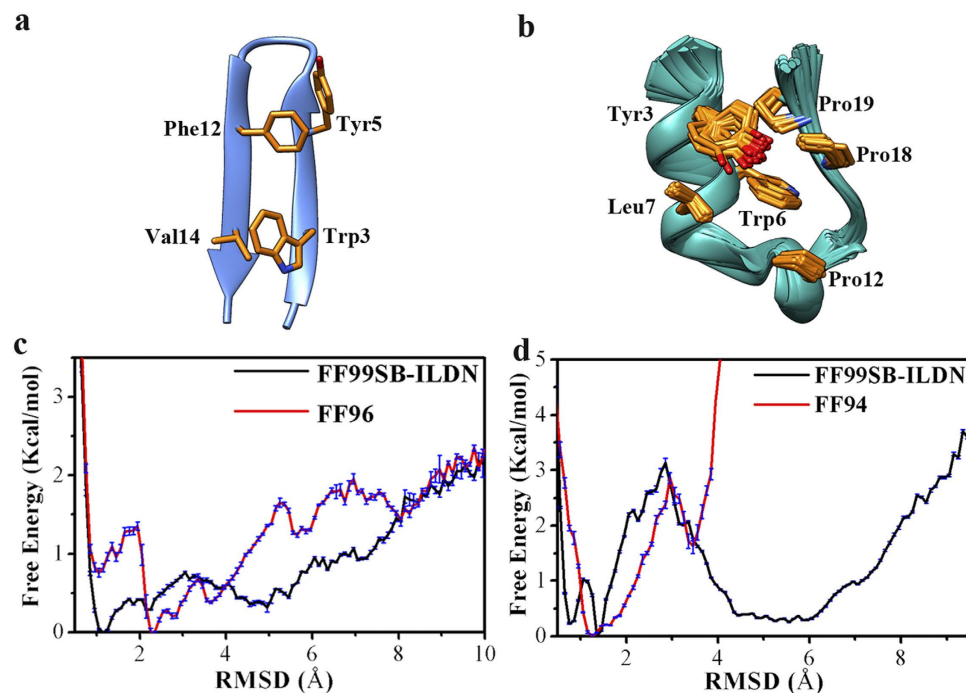
Yuqi Yu<sup>1</sup>, Jinan Wang<sup>1</sup>, Qiang Shao<sup>1</sup>, Jiye Shi<sup>2</sup> & Weiliang Zhu<sup>1</sup>

Protein folding is subject to the effects of solvation environment. A variety of organic solvents are used as additives for *in vitro* refolding of denatured proteins. Examination of the solvent effects on protein folding could be of fundamental importance to understand the molecular interactions in determining protein structure. This article investigated the folding of  $\alpha$ -helix and  $\beta$ -hairpin structures in water and the solutions of two representative refolding additives (methanol (MeOH) and 1-Ethyl-3-methylimidazolium chloride (EMIM-Cl) ionic liquid) using REMD simulations. For both  $\alpha$ -helix and  $\beta$ -hairpin in MeOH/water solution or  $\alpha$ -helix in EMIM-Cl/water solution, the transient structures along the folding pathway are consistent with the counterparts in water but the relative statistical weights are changed, leading to the decrease in the overall folding free energy barrier. Accordingly, MeOH promotes the folding of both  $\alpha$ -helix and  $\beta$ -hairpin but EMIM-Cl ionic liquid only promotes the folding of  $\alpha$ -helix, consistent with experimental observations. The present study reveals for the first time the trivial effects on folding route but significant effects on folding thermodynamics from MeOH and EMIM-Cl, explaining the function of protein refolding additives and testifying the validity of the folding mechanism revealed by *in vitro* protein folding study using refolding additives.

Protein folding is a molecular self-assembly process in which a disordered polypeptide collapses to form a well-defined three-dimensional biologically functional structure. Understanding the molecular mechanism underlying protein folding is of fundamental importance to biology. A widely held view is that the evolution has selected a protein's sequence in which the molecular interactions in native structure are mutually supportive and cooperatively lead to the functional structure, meaning that a given sequence is only consistent with a single native structure. The native structure of protein, as stated in free energy theory<sup>1,2</sup>, has the lowest energy in the vast protein conformational space. In this scenario, the energy landscape of protein folding is funnel-shaped and biased towards a single attractive basin of the native structure<sup>1,2</sup>. Any "trapping" protein conformations along the folding pathway have energetic depth smaller than the overall bias to the native structure, which guarantees that the native structure is both thermodynamically favorable and kinetically accessible.

Besides intrinsic properties of intra-protein interactions, protein folding is affected by extrinsic factors such as solvation environment as well. Any change in solvent condition might lead to the changes in several key aspects of protein stability and folding<sup>3,4</sup>. A conspicuous example is that water-soluble proteins can still fold into their native structures in specific organic solvents but the folding rates vary depending on the organic solvents used for protein solvation. Accordingly, protein folding is often investigated through *in vitro* refolding of denatured protein structure in various organic solvent solutions instead of pure water<sup>5,6</sup>. While the changes in thermodynamic and kinetic aspects of protein folding induced by organic solvents can be measured using the currently available experimental techniques, it is still a great challenge to monitor the folding process of protein in real time at the level of a single protein molecule. As a result, whether the folding route of protein is changed as a result of the change in solvation environment remains elusive, which may call into question the validity of the folding mechanism of protein revealed by *in vitro* experiments using organic solvents as refolding additives. A systematic

<sup>1</sup>Drug Discovery and Design Center, CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai, 201203, China. <sup>2</sup>UCB Biopharma SPRL, Chemin du Foriest, Braine-l'Alleud, Belgium. Correspondence and requests for materials should be addressed to Q.S. (email: qshao@mail.shnc.ac.cn) or J.S. (email: Jiye.Shi@ucb.com) or W.Z. (email: wlzhu@mail.shnc.ac.cn)



**Figure 1.** (a,b) NMR structures of GB1p and Trp-cage (the side-chains involved in the hydrophobic core cluster of protein are shown with yellow colored licorice representation). (c) One-dimensional free energy profile at 300 K as the function of the backbone RMSD with respect to the experimentally determined folded structure for GB1p in water measured by FF99SB-ILDN and FF96 force fields. (d) One-dimensional free energy profile at 300 K as the function of the backbone RMSD with respect to the experimentally determined folded structure for Trp-cage in water measured by FF99SB-ILDN and FF94 force fields.

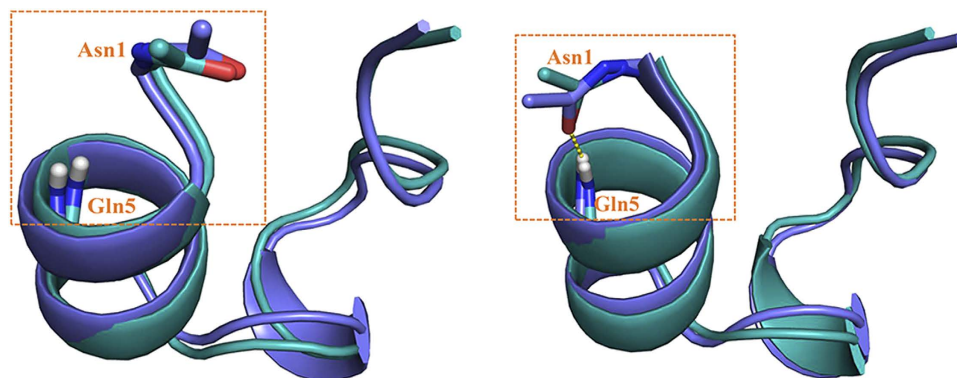
investigation of the effects of organic solvents (working as refolding additives) on not only folding thermodynamics/kinetics but also detailed folding route becomes much necessary.

Alcohols such as methanol (MeOH) and trifluoroethanol (TFE) are typical organic solvents which enhances the stabilization of  $\alpha$ -helical and  $\beta$ -hairpin secondary structures but meanwhile destabilizes the tertiary structure, generating a “molten globule” like state (a common folding intermediate state for small globular proteins)<sup>5,7</sup>. Multiple experiments and simulations<sup>8–12</sup> further showed that MeOH and TFE promote the *in vitro* refolding of denatured  $\alpha$ -helices and  $\beta$ -structures as well. Ionic liquids (ILs), molten salts usually composed of large-size organic cations (e.g., alkylsubstituted imidazolium, ammonium) and compact counterions, are another representative “organic solvent” which exert significant influence on protein folding, protein structural stability, and enzymatic activity. A great number of experiments have indicated that given suitable anions (e.g.,  $\text{BF}_4^-$ ,  $\text{PF}_6^-$ , and  $\text{Tf}_2\text{N}^-$ ), ILs stabilize the native structures of proteins<sup>13–17</sup>, promote the refolding process<sup>6,18,19</sup>, and improve the catalytic activity of enzymes of widely diverging types<sup>20–27</sup>. Examination of the effects of abovementioned organic solvents on protein folding can serve as a useful model system for understanding the various molecular interactions in determining protein structures.

$\alpha$ -Helices and  $\beta$ -hairpins are the most common structural motifs in protein and the detailed study could shed lights on the mechanism of protein folding. A typical  $\beta$ -hairpin is characterized by a hydrophobic core cluster packed by hydrophobic side-chains from anti-parallel strands as well as the backbone hydrogen bond assembly along the strands. This complex balance of local and nonlocal interactions makes  $\beta$ -hairpins resemble the folding of globular proteins, and the relevant study provides information about early events in protein folding. Therefore, short polypeptides adopting either  $\alpha$ -helix or  $\beta$ -hairpin or the combination of both structural motifs have attracted great attention in the last several decades. In the present study, using Trp-cage and the C-terminal  $\beta$ -hairpin (residues 41–56) from the B1 domain of protein G (GB1p) as protein models, we ran explicit solvent replica-exchange molecular dynamics (REMD) simulations to investigate atomically the folding of  $\alpha$ -helix and  $\beta$ -hairpin structures in not only water but also ~45% (v/v) MeOH/water solution and ~3M 1-Ethyl-3-methylimidazolium chloride (EMIM-Cl) ionic liquid solution. The accuracy of multiple physics based force fields was first assessed in the simulations of proteins in water to determine the force field with best balance of the  $\alpha$ -helix and  $\beta$ -hairpin tendencies. The structural characterization and energetics of the folding of Trp-cage and GB1p were then measured in all three abovementioned solutions, which provides comprehensive, atomic-level picture of the solvent effects on protein folding pathway as well as the associated folding thermodynamics.

## Results

**Force field assessment for the folding of  $\alpha$ -helix and  $\beta$ -hairpin structures in water.** A variety of force fields have been developed to measure the structural and dynamic properties of protein. Bias favoring



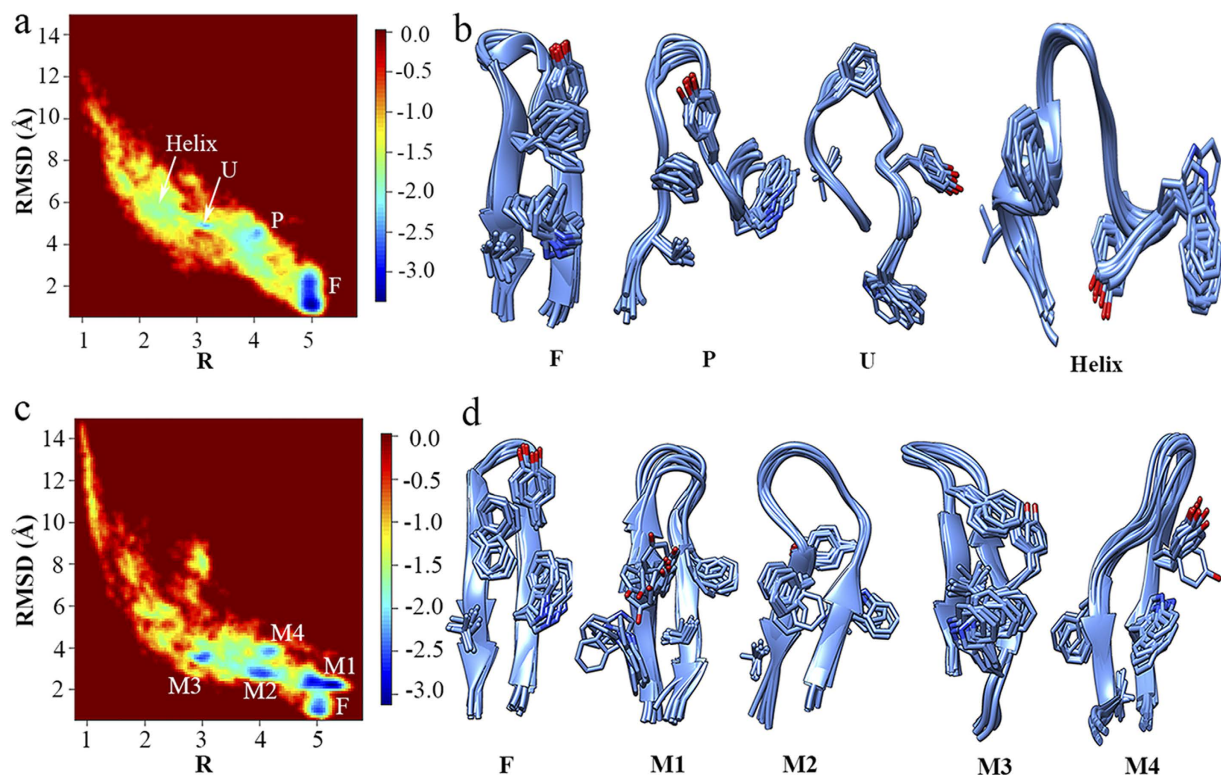
**Figure 2.** Comparison of the representative conformations (cyan colored) corresponding to the two local minima located at the low RMSD region in the free energy profile generated by FF99SB-ILDN force field (Fig. 1d) and the representative conformations of NMR structure ensembles of Trp-cage (blue colored). N-terminal backbone hydrogen bond (Asn1O-Gln5H) is highlighted with yellow square.

either  $\alpha$ -helical or  $\beta$ -hairpin secondary structure, however, has been observed among the force fields. The use of different force fields may produce different folding-related properties of protein. Multiple literatures have critically evaluated the performance of commonly used force fields in protein folding simulations, focusing on how accurately they reproduce the native structure and the dynamics of proteins, the experimental quantities relevant to protein folding<sup>28–31</sup>. In the present study, we chose several widely used AMBER force fields (namely AMBER FF94, FF96, and FF99SB-ILDN) to simulate the folding of GB1p and Trp-cage polypeptides in water solution. The most suitable force field for both  $\alpha$ -helix and  $\beta$ -hairpin structures was determined through the comparison of the conformational landscapes, which was then used for the subsequent investigation of solvent effects on protein folding.

In REMD simulation, random walk in temperature space is essential for replicas to escape local minima so as to make a sufficient sampling for protein configuration space. For each simulation system under study, the representative replica is ergodic for every temperature in the desired temperature range (from 300 K to 550 K) during the simulation, showing the efficiency of REMD search (see Supplementary Figs S1 and S2 online). In addition, as shown in Supplementary Figs S3 and S4 online, the fraction of folded structure at room-temperature (300 K) keeps increasing from the very beginning of each simulation and reaches a maximum plateau eventually. Taking the simulation systems of GB1p and Trp-cage in water as examples, we calculated the time series of the fraction of all populated structures of GB1p and Trp-cage at 300 K. As shown in Supplementary Figs S5 and S6 online, all the fractions reach the plateau. All the information (Figs S1–S6) indicates the convergence of REMD simulations in the present study. It is noteworthy that under FF99SB-ILDN force field, the folded  $\beta$ -hairpin structure of GB1p is most populated in the equilibrium conformational ensemble in water and its population ( $\sim 36\%$  at 300 K) (Figs S3a and S5a) is very close to the experimentally measured population of  $\sim 45\%$  by Munoz *et al.*<sup>32</sup> and  $30\%$  by Fesinmeyer *et al.*<sup>33</sup> On the other hand, the population of Trp-cage folded structure in water simulated by FF99SB-ILDN force field is consistent with the counterpart generated by the recent extensive conventional MD simulation of Lindorff-Larsen *et al.*<sup>34</sup> using CHARMM22\* force field. Both simulated population of Trp-cage are, however, relatively lower than the experimental data<sup>35,36</sup>. The discrepancy between simulations and experiments might be attributed to the small residual errors in the force field parameterization of two plentiful residues of proline and glycine in Trp-cage, which have an unusually large effect on the stability of this fold<sup>34</sup>.

Using backbone root-mean-square deviation (RMSD) with respect to the experimentally determined native structure as the reaction coordinate, the one-dimensional free energy profiles were depicted in Fig. 1 for the folding of GB1p and Trp-cage in water. The free energy landscape was calculated with the normalized probability,  $P(x) = Z^{-1}e^{-\beta W(x)}$  from a histogram analysis, where  $x$  is any set of reaction coordinate(s).  $W(x_2) - W(x_1) = -\frac{1}{\beta} \ln \frac{P(x_2)}{P(x_1)}$  is the relative free energy or potential of mean force. It is noteworthy that all free energy profiles including Fig. 1 and the others hereinafter were calculated on the basis of the simulation data at 300 K. Accordingly, the investigation of protein folding focuses on the temperature of 300 K. For GB1p (Fig. 1c), the FF99SB-ILDN force field generates a lowest-energy basin located at the smallest RMSD ( $\sim 1.0$  Å) whereas the FF96 force field makes the lowest-energy basin deviate to larger RMSD ( $\sim 2.2$  Å). Therefore, within the similar simulation time, FF99SB-ILDN force field is better than FF96 at constructing the conformational landscape of GB1p in which the native structure is supposed to have the lowest free energy level.

Meanwhile, the folding free energy profile of Trp-cage in water under FF99SB-ILDN force field (Fig. 1d) has the same shape as that generated by previous MD simulation of Lindorff-Larsen *et al.*<sup>34</sup>: the local minima corresponding to the folded and unfolded states are separated by a transition state (at RMSD  $\approx 2.8$  Å), implying that the folding of Trp-cage is a two-state transition (see the comparison of the folding free energy profiles in Supplementary Fig. S7 online). Ensembles of native structures of Trp-cage have been identified by previous NMR experiment, with the difference mainly at the N-terminal region: the N-terminus is flexible and the N-terminal backbone hydrogen bond (Asn1O-Gln5H) is either formed or broken<sup>37</sup>. Intriguingly, the low RMSD region ( $< 2.0$  Å) in the free energy profile generated by FF99SB-ILDN (Fig. 1d) is separated into two local minima. The

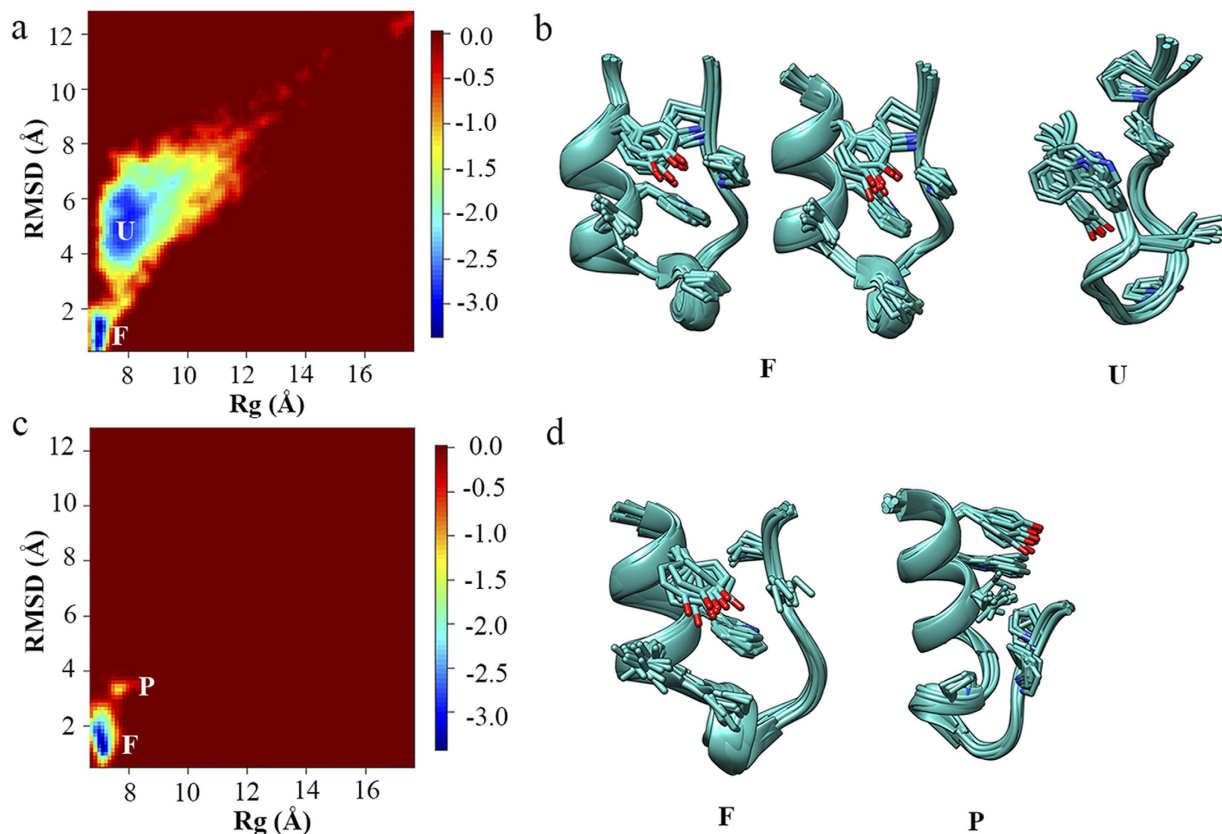


**Figure 3.** Two-dimensional free energy landscape as the function of the backbone RMSD and “R” parameter for the folding of GB1p in water and the representative conformations for all distinct states in the free energy landscape measured by (a,b) FF99SB-ILDN and (c,d) FF96 force fields.

representative conformations of the corresponding native-like structures are close to each other except that the N-terminal backbone hydrogen bond is formed in one conformation but broken in the other, perfectly reproducing the experimentally observed trivial difference among structure ensembles of Trp-cage in water (Fig. 2). On the other hand, while the free energy profile of Trp-cage generated by FF94 force field also reflects a two-state transition of protein folding, the free energy barrier from the unfolded to folded state is significantly smaller than the counterpart under FF99SB-ILDN. In addition, the free energy profile under FF94 is limited in much narrower RMSD range than the free energy profile under FF99SB-ILDN. All the information suggests that the former force field has greater bias favoring  $\alpha$ -helix structure than the latter.

To further assess the force fields in describing the folding of  $\beta$ -hairpin structure in water, two-dimensional free energy landscape was projected onto the backbone RMSD and a robust “R” parameter ( $R = \sum_{i=1}^n \frac{r_i^0}{r_i}$ , where  $r_i^0$  is the inter-strand  $C_\alpha$ - $C_\alpha$  distance in the native structure and  $r_i$  is the same distance in the simulation trajectory<sup>38,39</sup>). “R” parameter was chosen because of its ability of reflecting the formation status of the key structural elements in the folding process of  $\beta$ -hairpin and thus mapping the folding pathway in conformational landscape. As a comparison, radius of gyration ( $R_g$ ) was also used in combination with RMSD to draw two-dimensional free energy landscapes for GB1p in all solutions under study. The distinct states which can be clearly indicated in individual free energy landscapes of RMSD and R are undistinguished in the corresponding free energy landscapes of RMSD and  $R_g$  (see Supplementary Fig. S8 online). Hierarchical clustering analyses were run using *kclust* algorithm available in MMTSM Toolset<sup>40</sup> to identify the representative (most populated) conformations for all distinct states indicated in Fig. 3. Four distinct states are presented in the folding free energy landscape of GB1p in water simulated by FF99SB-ILDN force field (Fig. 3a,b): the folded (F) state containing all native structural elements, the partially folded (P) state in which only the two turn-neighboring hydrophobic side-chains (Tyr5 and Phe12) are packed and the surrounding backbone hydrogen bonds are formed, the unfolded (U) state with relatively extended structure, and a state adopting mainly  $\alpha$ -helical structure which is thus defined as Helix state here. The connection of these distinct states maps the folding pathway of GB1p. The free energy landscape of GB1p in water simulated by FF96 force field (Fig. 3c,d) is, however, much more complex: besides the folded (F) state, multiple misfolded  $\beta$ -hairpin structures (M1-M4) also exist (see the representative conformations in Fig. 3d and the corresponding backbone hydrogen bond assembly for all states in Supplementary Table S3 online). The formation of multiple  $\beta$ -hairpin structures could be attributed to the strong  $\beta$ -structure bias of FF96 force field.

The two-dimensional free energy landscape was projected onto the backbone RMSD and  $R_g$  for Trp-cage in water to show the force field effects on the folding of  $\alpha$ -helix (“R” parameter is not suitable for  $\alpha$ -helix structure). As shown in Fig. 4, two distinct states corresponding to the folded (F) and unfolded (U) states are presented in the free energy landscape simulated by FF99SB-ILDN force field. The representative conformations of the former

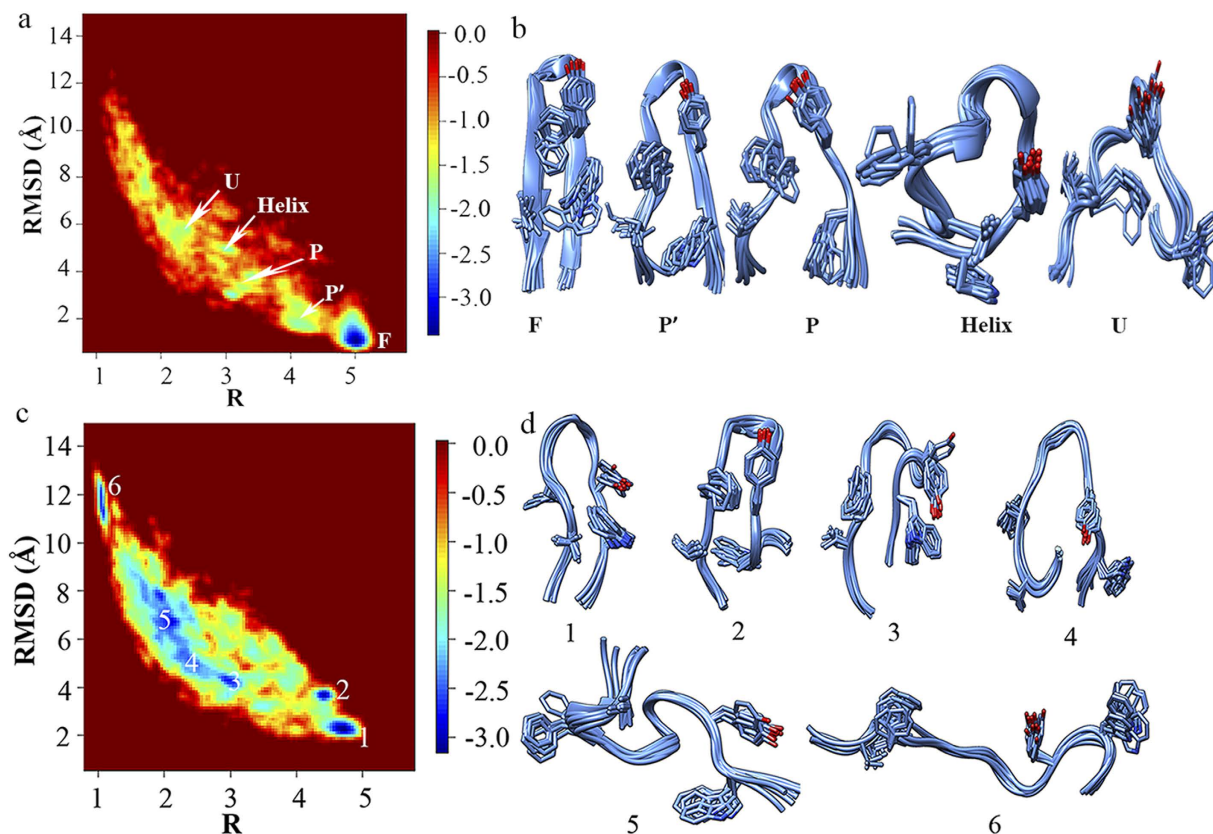


**Figure 4.** Two-dimensional free energy landscape as the function of the backbone RMSD and  $R_g$  for the folding of Trp-cage in water and the representative conformations for all distinct states in the free energy landscape measured by (a,b) FF99SB-ILDN and (c,d) FF94 force fields.

state are consistent with the NMR experimentally measured native structure ensembles (Fig. 2). In contrast, no secondary structure element is formed in the representative conformation of the latter state (Fig. 4b). On the other hand, the  $\alpha$ -helix structure is well formed in both distinct states in the free energy landscape simulated by FF94 force field, making one state as the folded state and the other more likely as the partially folded (P) state, which, again, indicates strong  $\alpha$ -helix bias of FF94 force field. In summary, among all force fields under study, FF99SB-ILDN is the most suitable force field for mimicking the folding of both  $\alpha$ -helix and  $\beta$ -hairpin structures.

**Methanol promoting the folding of both  $\alpha$ -helix and  $\beta$ -hairpin Structures.** Next, the folding of GB1p and Trp-cage in  $\sim 45\%$  (v/v) MeOH/water and 3M EMIM-Cl/water solutions was simulated using FF99SB-ILDN force field and REMD methodology, respectively. The one-dimensional free energy profile (see Supplementary Fig. S9 online) indicates that the folded state of GB1p in MeOH/water solution has the lowest energy level along the reaction coordinate. The transient states in the folding pathway of GB1p in MeOH/water have higher free energy level (relative to the folded state) than the counterparts of GB1p in water. As a result, the folding free energy barrier of GB1p ( $\Delta G_f^\ddagger$ ) is  $\sim 0.27$  kcal/mol in MeOH/water solution, smaller than the value of 0.40 kcal/mol in water, suggesting that methanol promotes the folding of  $\beta$ -hairpin structure by decreasing the folding free energy barrier. This result is consistent with the observations in multiple experiments. For instance, using circular dichroism (CD) and NMR spectra, Searle *et al.* found that the addition of methanol or TFE promotes the native-like structure formation of the N-terminal hairpin of ferredoxin I which adopts dominantly random coil structure in pure water<sup>9</sup>. In addition, the NMR experiment by Platt *et al.*<sup>41</sup> observed that the folding of the  $\beta$ -hairpin structure of a mutant of yeast ubiquitin is accelerated in low-concentrated methanol and particularly TFE aqueous solutions.

The four distinct states (F, P, U, and Helix states) which are presented in the two-dimensional folding free energy landscape of GB1p in water (Fig. 3a,b) can be also seen in the free energy landscape of GB1p in MeOH/water solution (Fig. 5a,b). The representative conformations of abovementioned states in MeOH/water solution are superposed onto the counterparts in water, respectively (Fig. 6). One can see that the structure of each state is largely consistent in the two solutions (e.g., the F state has only trivial difference in the orientation of Trp3 side-chain, the P state has the structure difference only in the flapping terminal regions, the U and Helix states have the structure difference mainly in the central region). In addition, another partially folded (P') state is also presented, which has structural feature even closer to the native state than the P state. The presence of the P' state might be attributed to methanol induced stabilization of native-like structure of  $\beta$ -hairpin.



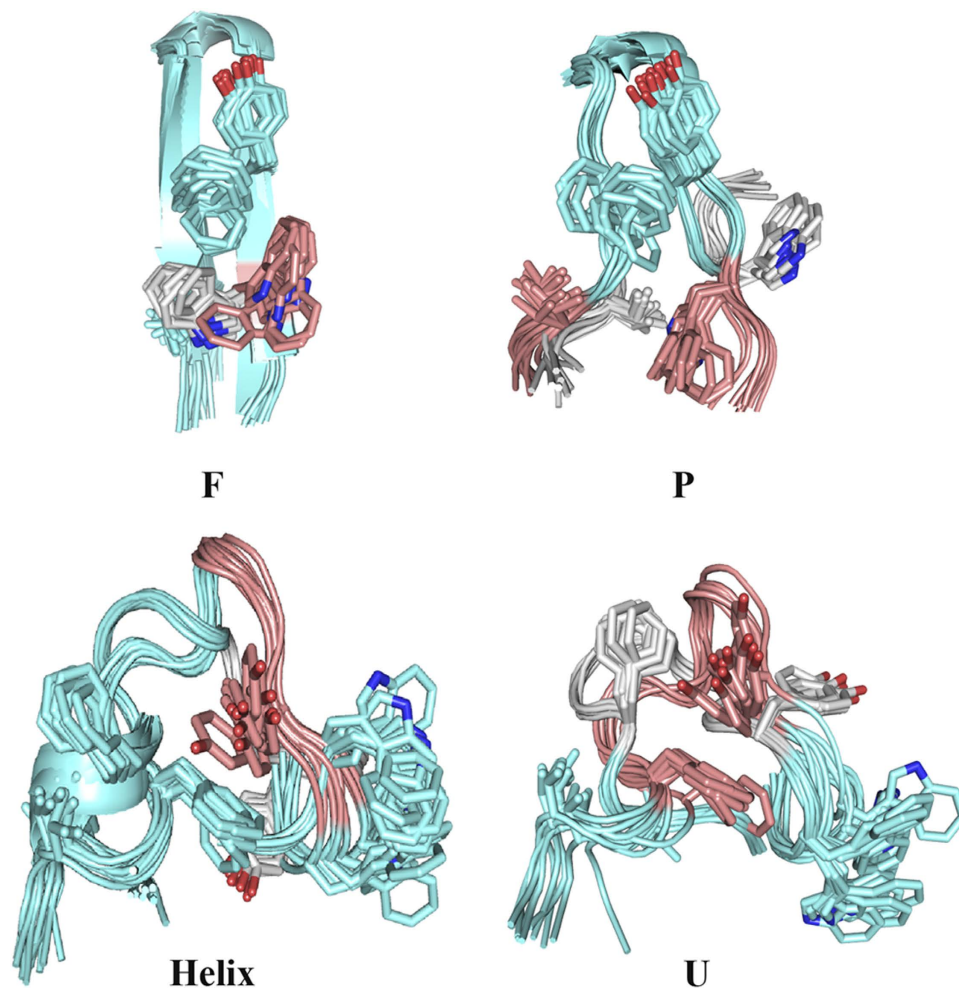
**Figure 5.** (a,b) Two-dimensional free energy landscape as the function of the backbone RMSD and “R” parameter for the folding of GB1p in MeOH/water solution measured by FF99SB-ILDN force field and the representative conformations for all distinct states in the free energy landscape. (c,d) Two-dimensional free energy landscape as the function of the backbone RMSD and “R” parameter for the folding of GB1p in EMIM-Cl/water solution measured by FF99SB-ILDN force field and the representative conformations for all distinct states in the free energy landscape.

On the other hand, the two-dimensional free energy landscape for Trp-cage in MeOH/water solution (Fig. 7a) has similar feature as that for Trp-cage in water except that the folding free energy barrier between the folded and unfolded states is smaller ( $\Delta G_f^\ddagger$  is 2.44 kcal/mol in MeOH/water solution and 2.66 kcal/mol in water), suggesting that methanol could promote the folding of  $\alpha$ -helix structure as well. The representative conformation of the unfolded state in MeOH/water is similar to that in water except that the N-terminal region is pointed to opposite direction relative to the C-terminus (see Supplementary Fig. S10 online). Interestingly, the representative conformation of the folded state of Trp-cage in MeOH/water has the N-terminal backbone hydrogen bond well formed, indicating that methanol enhances the stability of native structure of Trp-cage.

**EMIM-Cl ionic liquid promoting the folding of  $\alpha$ -helix but not  $\beta$ -hairpin.** Multiple ionic liquids have been reported to effectively promote the *in vitro* refolding of denatured proteins. Buchfink *et al.* tested a series of EMIM<sup>+</sup> salts as protein refolding enhancers and observed that the recombinant plasminogen activator rPA has the highest refolding yield in 3 M EMIM-Cl ionic liquid solution<sup>6</sup>. The CD spectroscopy experiment by Huang *et al.* observed that the  $\alpha$ -helix structured short peptides AKA<sub>2</sub> and Trp-cage can fold into their native structures in neat ionic liquid of 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide (C<sub>4</sub>mpy-Tf<sub>2</sub>N) and the formed native structures persist to very high temperature (up to 96 °C) whereas the  $\beta$ -hairpin structured peptide TRPZIP4 is largely destabilized in the same ionic liquid medium<sup>42</sup>.

The two-dimensional free energy landscape of Fig. 5c indicates that various structures from fully extended to compact ones consist of the equilibrium conformational ensemble of GB1p in EMIM-Cl/water solution. Nevertheless, even the compact structure which has the lowest RMSD in the free energy landscape (state 1 in Fig. 5c,d) is deviated from the native  $\beta$ -hairpin structure of GB1p: although the native  $\beta$ -turn configuration is formed, neither the cross-strand hydrophobic core cluster is well packed nor the backbone hydrogen bond assembly along the anti-parallel strands is formed in state 1. As a result, GB1p cannot fold spontaneously into its exact native structure in EMIM-Cl ionic liquid solution.

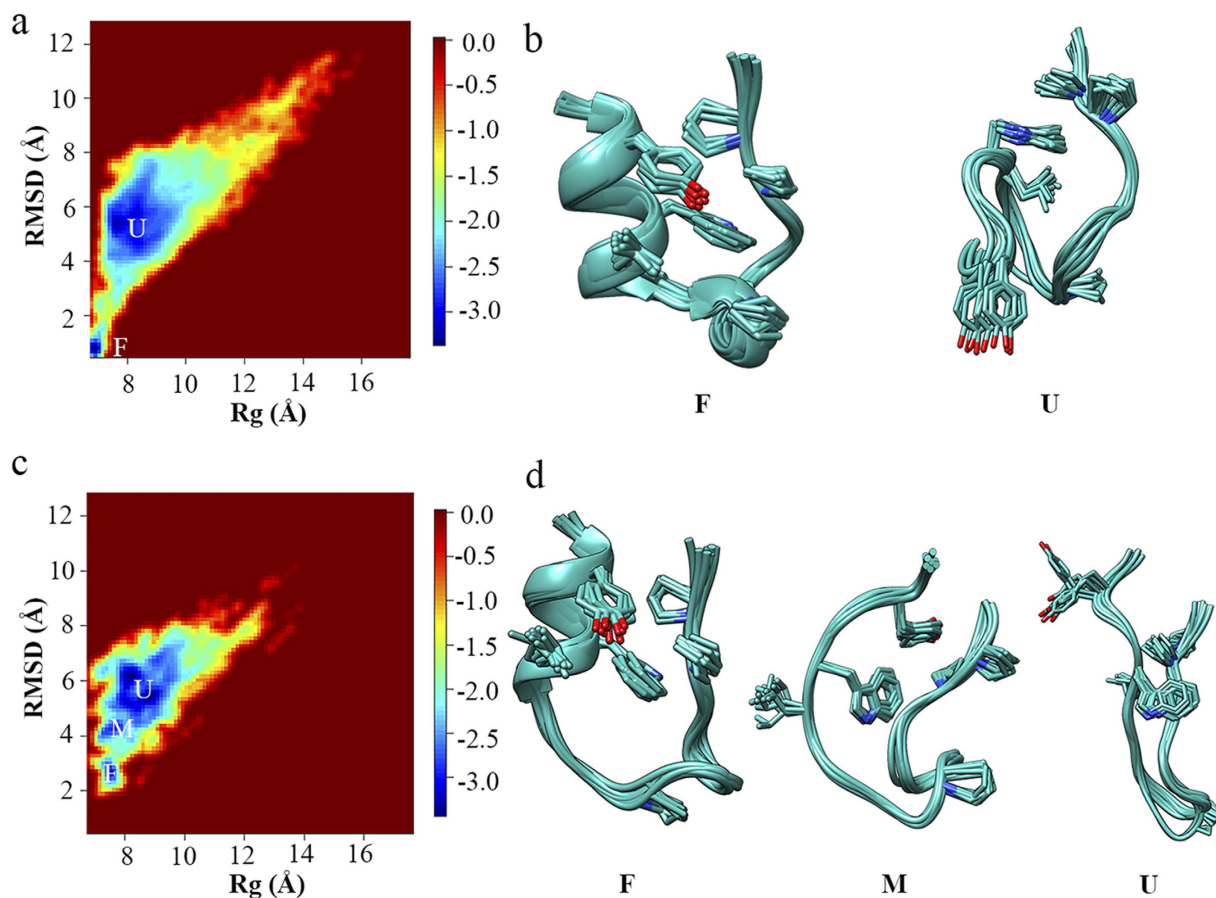
Unlike GB1p, Trp-cage can fold into its native-like structure in the same ionic liquid solution. In comparison to the native-like structures adopted in water and MeOH/water, the native-like structure adopted in EMIM-Cl/water solution is more deviated from the NMR experimentally measured native structure. Nevertheless, the key



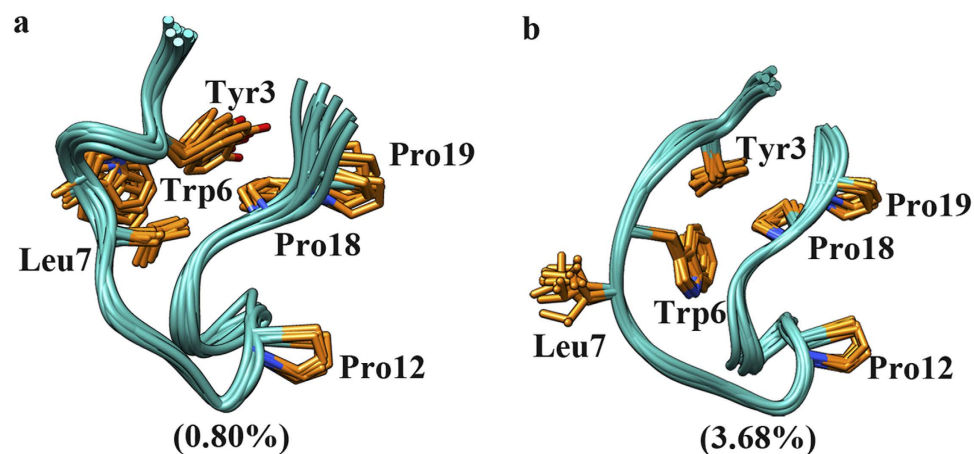
**Figure 6.** Comparison of the representative conformations of all distinct states of GB1p in MeOH/water solution and water. For each state, the regions of GB1p sharing similar conformations in two solutions are shown in green color whereas the regions having different structural features are shown in pink and grey colors in MeOH/water and water, respectively.

native structural elements of Trp-cage are still formed: the N-terminal helix is well folded, the C-terminal polyproline II region (including Pro18 and Pro19) is properly packed against Leu2, Tyr3, and Trp6 at the N-terminus (see the representative conformation of F state in Fig. 7d). The main deviation of the native-like structure in EMIM-Cl/water solution is that Pro12 within the  $3_{10}$ -helix (residues of Gly11-Ser14) is failed to pack with Trp6. In addition, the unfolded state of Trp-cage in EMIM-Cl/water solution has similar structure feature as the counterparts in water and MeOH/water solution. An additional state is presented between the folded and unfolded states, which contains no secondary structure content and is composed by mainly non-native intra-protein interactions as well as a small fraction of native hydrophobic contacts (Tyr3-Pro19 and Trp6-Pro18) (Fig. 8). As a result, this state is treated as a misfolded state here. The similar misfolded structure can be also seen in the equilibrium conformational ensembles in water and MeOH/water solution but with extremely low population. The overall folding free energy landscape of Trp-cage in the EMIM-Cl/water solution is flatter than that of Trp-cage in water and the folding free energy barrier from unfolded to folded state in the former case is lower than the latter ( $\Delta G_f^*$  is 1.65 kcal/mol in EMIM-Cl/water solution and 2.66 kcal/mol in water).

Combining Figs 3–8, one can see that methanol and EMIM-Cl ionic liquid have different influence on the folding of  $\alpha$ -helix and  $\beta$ -hairpin structures: while methanol promotes the folding of both  $\alpha$ -helix and  $\beta$ -hairpin structures, EMIM-Cl ionic liquid can only promote the folding of  $\alpha$ -helix but hinder the folding of  $\beta$ -hairpin indeed. In the case of the protein folding promoted by either solvent, the folding free energy barrier is decreased and meanwhile the detailed folding pathway is not significantly influenced. As shown in Fig. 9a, along the folding pathway of GB1p, the unfolded structure folds into a partially folded structure (or multiple partially folded structures in MeOH/water solution) in which the cross-strand hydrophobic core cluster is partially packed and the neighboring backbone hydrogen bonds are formed. The further folding of GB1p to the native  $\beta$ -hairpin structure includes the well packing of cross-strand hydrophobic core cluster and the complete formation of all backbone hydrogen bonds. The folding of  $\beta$ -hairpin thus consists of the sequential formation of native structural elements. In contrast, for  $\alpha$ -helical structure with a lower degree of cooperativity such as Trp-cage, the folding is a typical two-state transition. All native structural elements are formed in



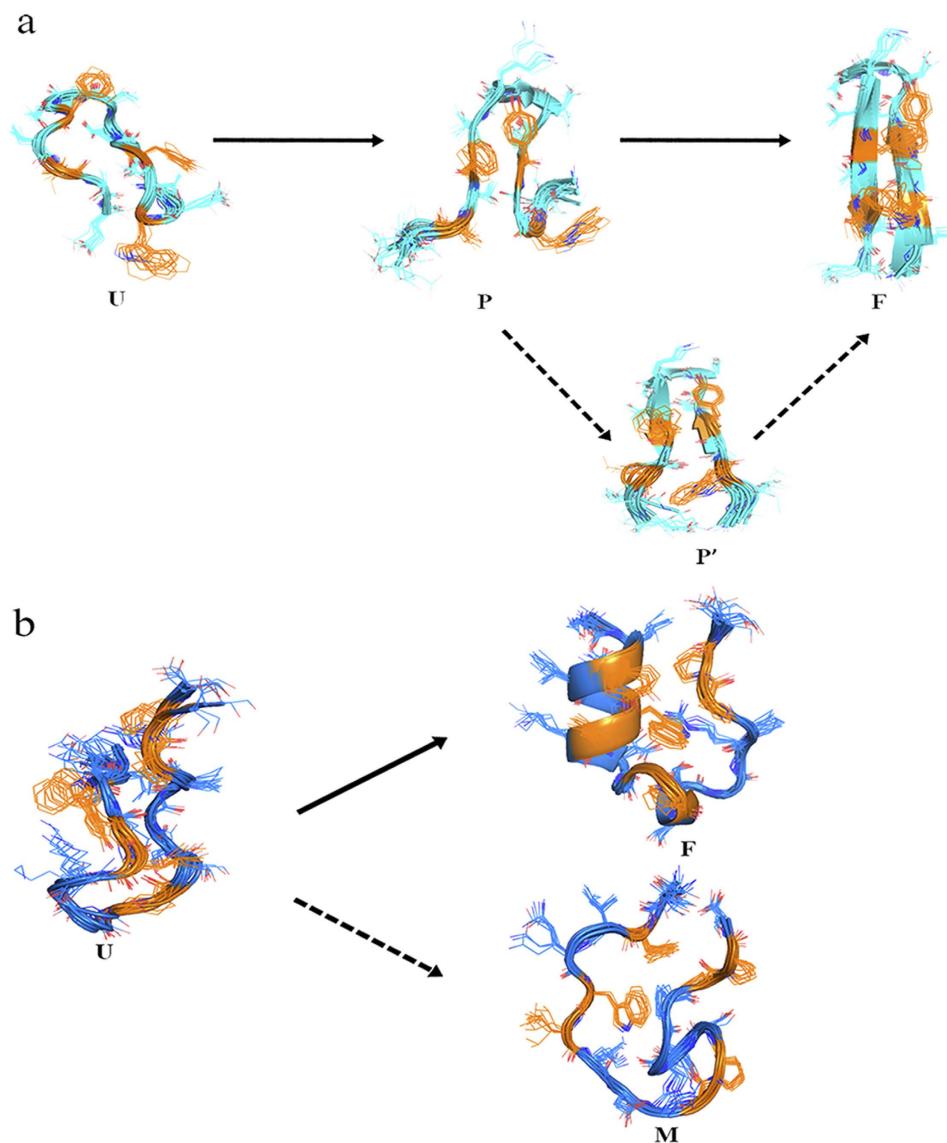
**Figure 7.** (a,b) Two-dimensional free energy landscape as the function of the backbone RMSD and  $R_g$  for the folding of Trp-cage in MeOH/water solution measured by FF99SB-ILDN force field and the representative conformations for all distinct states in the free energy landscape. (c,d) Two-dimensional free energy landscape as the function of the backbone RMSD and  $R_g$  for the folding of Trp-cage in EMIM-Cl/water solution measured by FF99SB-ILDN force field and the representative conformations for all distinct states in the free energy landscape.



**Figure 8.** Representative conformations of the misfolded state observed in the simulations of Trp-cage in (a) water and (b) EMIM-Cl ionic liquid solution, along with the population in the equilibrium conformational ensembles in the corresponding solutions.

single transition from the unfolded to folded states. An insignificant “misfolding” seems to hide behind experimentally detectable folding of Trp-cage and the presence of EMIM-Cl ionic liquid could slightly increase the probability of such misfolding.





**Figure 9.** Schematics of the folding pathway of (a) GB1p and (b) Trp-cage. For clarity, residues involved in the hydrophobic core cluster of protein are colored in yellow.

## Discussion

Protein folding is a consequence of complex inter-residue interactions, which should be mutually supportive and cooperatively lead to thermodynamically stable native structure<sup>2</sup>. A variety of experiments have indicated that the composition of the equilibrium conformational ensemble of protein folding is complex: not only the unfolded and folded states but also transiently populated “trapping” states (e.g., intermediate state) could exist<sup>43–46</sup>. By rapidly forming compact intermediate states, proteins could be driven along a preferred folding route toward their native structures and thus avoid time consuming long-range conformational search<sup>47</sup>. Discriminating the structural and energetic properties of various “trapping” conformations in the folding conformational ensemble sheds light on the molecular mechanism of protein folding.

Through a series of REMD simulations, the conformational landscapes of two protein models (namely GB1p and Trp-cage) were explored for their folding in water, ~45% MeOH/water, and ~3M EMIM-Cl/water solutions in the present study. The concentration of hydrated EMIM-Cl ionic liquid solution was set according to the experimental observation that EMIM-Cl is most efficient in promoting the refolding of denatured protein at 3M concentration<sup>6</sup>. The concentration of methanol in aqueous solution was chosen from the concentration range which has been often used in the experimental studies of methanol-induced protein conformational transition *in vitro*<sup>48,49</sup>. The relative statistical weights of various structures in the folding free energy landscape were properly estimated, providing comprehensive information of the folding mechanism of proteins in three solutions under study.

It is observed that methanol promotes the folding of both  $\alpha$ -helix and  $\beta$ -hairpin structures and meanwhile EMIM-Cl ionic liquid can only promote the folding of  $\alpha$ -helix but not  $\beta$ -hairpin, which are in great agreement

with the experimental observation for the two organic solvents<sup>6,8–12,18,19</sup>. The detailed analysis reveals that for both  $\alpha$ -helix and  $\beta$ -hairpin structures in MeOH/water solution or  $\alpha$ -helix in EMIM-Cl/water solution, the transient structures along the folding pathway is maintained as in pure water whereas the relative statistical weights are changed. As a result, the folding route is maintained but the overall folding free energy barrier is decreased and accordingly the folding of specific proteins is promoted in MeOH/water and EMIM-Cl/water solutions. Various organic solvents are used as refolding additives in *in vitro* protein folding experiments but whether the organic solvents change the folding route has been not investigated deeply. The present microscopic view of the protein folding in the presence of representative protein refolding additives (methanol and EMIM-Cl ionic liquid) reveals that organic solvents can have trivial effects on folding route but significant effects on folding thermodynamics, supporting the use of refolding additives in protein folding investigation.

## Methods

All-atom molecular dynamics simulations were performed using AMBER11 package. AMBER FF99SB-ILDN<sup>50</sup>, FF96<sup>51</sup>, and FF94<sup>52</sup> force fields were used to model protein atoms, TIP3P explicit solvent model<sup>53</sup> was used to mimic water molecules explicitly. In addition, the parameters of the force fields concerning methanol were taken from Caldwell and Kollman<sup>54</sup> and the parameters for EMIM<sup>+</sup> cation were taken from Canongia Lopes *et al.*<sup>55</sup> For each simulation system, a relatively extended structure of either Trp-cage or GB1p without any native structural element was placed into a cubic box containing a plenty of water (and MeOH or EMIM-Cl) molecules. Suitable counterions were then added into each system to balance the charge of protein. The detailed simulation parameters including the number of solvent molecules, box size, and simulation time for all simulation systems were organized in Supplementary Table S1 online.

For each system, REMD simulation<sup>56</sup> was run in the NVT ensemble. 38 ~ 40 replicas were used at temperatures ranging from 300 K to 550 K. The number of replicas were determined by the webserver (<http://folding.bmc.uu.se/remd/>)<sup>57</sup>, and the temperature distribution along the replicas was optimized using the algorithm of Nadker and Hansmann<sup>58</sup> to give a theoretical acceptance probability of 20%. The detailed temperature distribution for each system was listed in Supplementary Table S2 online. Exchanges between neighboring replicas were attempted every 1000 steps. The SHAKE algorithm<sup>59</sup> was applied to constrain the bonds connecting hydrogen atoms, allowing us to use 2.0 fs as the time step. Particle mesh Ewald (PME) was applied to handle the long-range electrostatic interactions and the non-bonded cutoff of 10 Å was used<sup>60</sup>. The Langevin dynamics with a collision frequency of 3.0 ps<sup>-1</sup> was adopted to control the temperature of the system. The coordinates were saved every 1000 steps. The simulation time per simulation system is accumulated to 2.0 ~ 6.0  $\mu$ s.

## References

1. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels - a kinetic approach to the sequence structure relationship. *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725 (1992).
2. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
3. Canchi, D. R. & Garcia, A. E. Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.* **64**, 273–293 (2013).
4. Mattos, C. & Ringe, D. Proteins in organic solvents. *Curr. Opin. Struct. Biol.* **11**, 761–764 (2001).
5. Buck, M. Trifluoroethanol and colleagues: Cosolvents come of age. Recent studies with peptides and proteins. *Q. Rev. Biophys.* **31**, 297–355 (1998).
6. Buchfink, R., Tischer, A., Patil, G., Rudolph, R. & Lange, C. Ionic liquids as refolding additives: Variation of the anion. *J. Biotechnol.* **150**, 64–72 (2010).
7. Uversky, V. N., Narizhneva, N. V., Kirschstein, S. O., Winter, S. & Lober, G. Conformational transitions provoked by organic solvents in  $\beta$ -lactoglobulin: Can a molten globule like intermediate be induced by the decrease in dielectric constant? *Fold. Des.* **2**, 163–172 (1997).
8. Blanco, F. J., Angeles Jimenez, M., Pineda, A., Rico, M., Santoro, J. & Nieto, J. L. NMR solution structure of the isolated N-terminal fragment of protein-G B1 domain. Evidence of trifluoroethanol induced native-like  $\beta$ -hairpin formation. *Biochemistry* **33**, 6004–6014 (1994).
9. Searle, M. S., Zerella, R., Dudley, D. H. & Packman, L. C. Native-like  $\beta$ -hairpin structure in an isolated fragment from ferredoxin: NMR and CD studies of solvent effects on the N-terminal 20 residues. *Protein Eng.* **9**, 559–565 (1996).
10. Harding, M. M., Williams, D. H. & Woolfson, D. N. Characterization of a partially denatured state of a protein by 2-dimensional NMR - reduction of the hydrophobic interactions in ubiquitin. *Biochemistry* **30**, 3120–3128 (1991).
11. Blanco, F. J., Rivas, G. & Serrano, L. A short linear peptide that folds into a native stable  $\beta$ -hairpin in aqueous solution. *Nat. Struct. Biol.* **1**, 584–590 (1994).
12. Roccatano, D., Colombo, G., Fioroni, M. & Mark, A. E. Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study. *Proc. Natl. Acad. Sci. USA* **99**, 12179–12184 (2002).
13. De Diego, T., Lozano, P., Gmouh, S., Vaultier, M. & Iborra, J. L. Fluorescence and CD spectroscopic analysis of the  $\alpha$ -chymotrypsin stabilization by the ionic liquid, 1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide. *Biotechnol. Bioeng.* **88**, 916–924 (2004).
14. Fujita, K., MacFarlane, D. R. & Forsyth, M. Protein solubilising and stabilising ionic liquids. *Chem. Commun.* doi: 10.1039/B508238B (2005).
15. Fujita, K., Forsyth, M., MacFarlane, D. R., Reid, R. W. & Elliott, G. D. Unexpected improvement in stability and utility of cytochrome c by solution in biocompatible ionic liquids. *Biotechnol. Bioeng.* **94**, 1209–1213 (2006).
16. Weingartner, H., Cabrele, C. & Herrmann, C. How ionic liquids can help to stabilize native proteins. *Phys. Chem. Chem. Phys.* **14**, 415–426 (2012).
17. Lozano, P., De Diego, T., Carrie, D., Vaultier, M. & Iborra, J. L. Over-stabilization of *Candida antarctica* lipase B by ionic liquids in ester synthesis. *Biotechnol. Lett.* **23**, 1529–1533 (2001).
18. Summers, C. A. & Flowers, R. A. Protein renaturation by the liquid organic salt ethylammonium nitrate. *Protein Sci.* **9**, 2001–2008 (2000).
19. Lange, C., Patil, G. & Rudolph, R. Ionic liquids as refolding additives: N<sup>-</sup>-alkyl and N<sup>-</sup>-( $\omega$ -hydroxyalkyl) N-methylimidazolium chlorides. *Protein Sci.* **14**, 2693–2701 (2005).
20. Kragl, U., Eckstein, M. & Kaftzik, N. Enzyme catalysis in ionic liquids. *Curr. Opin. Biotech.* **13**, 565–571 (2002).
21. van Rantwijk, F., Lau, R. M. & Sheldon, R. A. Biocatalytic transformations in ionic liquids. *Trends Biotechnol.* **21**, 131–138 (2003).
22. Yang, Z. Hofmeister effects: An explanation for the impact of ionic liquids on biocatalysis. *J. Biotechnol.* **144**, 12–22 (2009).

23. Sheldon, R. A., Lau, R. M., Sorgedraeger, M. J., van Rantwijk, F. & Seddon, K. R. Biocatalysis in ionic liquids. *Green Chem.* **4**, 147–151 (2002).
24. Schofer, S. H., Kaftzik, N., Wasserscheid, P. & Kragl, U. Enzyme catalysis in ionic liquids: lipase catalysed kinetic resolution of 1-phenylethanol with improved enantioselectivity. *Chem. Commun.* doi: 10.1039/B009389K (2001).
25. Park, S. & Kazlauskas, R. J. Improved preparation and use of room-temperature ionic liquids in lipase-catalyzed enantio- and regioselective acylations. *J. Org. Chem.* **66**, 8395–8401 (2001).
26. Berberich, J. A., Kaar, J. L. & Russell, A. J. Use of salt hydrate pairs to control water activity for enzyme catalysis in ionic liquids. *Biotechnol. Progr.* **19**, 1029–1032 (2003).
27. Yang, Z., Yue, Y. J. & Xing, M. Tyrosinase activity in ionic liquids. *Biotechnol. Lett.* **30**, 153–158 (2008).
28. Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **24**, 98–105 (2014).
29. Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Systematic validation of protein force fields against experimental data. *Plos One* **7**, e32131 (2012).
30. Lange, O. F., van der Spoel, D. & de Groot, B. L. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.* **99**, 647–655 (2010).
31. Beauchamp, K. A., Lin, Y. S., Das, R. & Pande, V. S. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* **8**, 1409–1414 (2012).
32. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of  $\beta$ -hairpin formation. *Nature* **390**, 196–199 (1997).
33. Fesinmeyer, R. M., Hudson, F. M. & Andersen, N. H. Enhanced hairpin stability through loop design: The case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* **126**, 7238–7243 (2004).
34. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **334**, 517–520 (2011).
35. Barua, B., Lin, J. C., Williams, V. D., Kummeler, P., Neidigh, J. W. & Andersen, N. H. The Trp-cage: optimizing the stability of a globular miniprotein. *Protein Eng. Des. Sel.* **21**, 171–185 (2008).
36. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9**, 425–430 (2002).
37. Struthers, M., Ottesen, J. J. & Imperiali, B. Design and NMR analyses of compact, independently folded BBA motifs. *Fold. Des.* **3**, 95–103 (1998).
38. Yang, S. C., Onuchic, J. N., Garcia, A. E. & Levine, H. Folding time predictions from all-atom replica exchange simulations. *J. Mol. Biol.* **372**, 756–763 (2007).
39. Thukral, L., Smith, J. C. & Daidone, I. Common folding mechanism of a  $\beta$ -hairpin peptide via non-native turn formation revealed by unbiased molecular dynamics simulations. *J. Am. Chem. Soc.* **131**, 18147–18152 (2009).
40. Feig, M., Karanicolas, J. & Brooks, C. L., 3rd. MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **22**, 377–395 (2004).
41. Platt, G. W., Simpson, S. A., Layfield, R. & Searle, M. S. Stability and folding kinetics of a ubiquitin mutant with a strong propensity for nonnative  $\beta$ -hairpin conformation in the unfolded state. *Biochemistry* **42**, 13762–13771 (2003).
42. Huang, J. L., Noss, M. E., Schmidt, K. M., Murray, L. & Bunagan, M. R. The effect of neat ionic liquid on the folding of short peptides. *Chem. Commun.* **47**, 8007–8009 (2011).
43. Nguyen, H., Jager, M., Moretto, A., Gruebele, M. & Kelly, J. W. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA* **100**, 3948–3953 (2003).
44. Gsponer, J. *et al.* Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. USA* **103**, 99–104 (2006).
45. Hu, K. N., Yau, W. M. & Tycko, R. Detection of a transient intermediate in a rapid protein folding process by solid-state nuclear magnetic resonance. *J. Am. Chem. Soc.* **132**, 24–25 (2010).
46. Vallee-Belisle, A. & Michnick, S. W. Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat. Struct. Mol. Biol.* **19**, 731–736 (2012).
47. Roder, H. & Colon, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28 (1997).
48. Hirota, N., Mizuno, K. & Goto, Y. Group additive contributions to the alcohol-induced  $\alpha$ -helix formation of melittin: Implication for the mechanism of the alcohol effects on proteins. *J. Mol. Biol.* **275**, 365–378 (1998).
49. Yamazaki, K., Iwura, T., Ishikawa, R. & Ozaki, Y. Methanol-induced, tertiary and secondary structure changes of granulocyte-colony stimulating factor. *J. Biochem.* **140**, 49–56 (2006).
50. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. & Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
51. Kollman, P. A., Dixon, R., Cornell, W., Fox, T., Chipot, C. & Pohorille, A. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. *Computer Simulation of Biomolecular Systems* vol. **3**, 83–96 (1997).
52. Cornell, W. D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
54. Caldwell, J. W. & Kollman, P. A. Structure and properties of neat liquids using nonadditive molecular dynamics: Water, methanol, and N-methylacetamide. *J. Phys. Chem.* **99**, 6208–6219 (1995).
55. Lopes, J. N. C., Deschamps, J. & Padua, A. A. H. Modeling ionic liquids using a systematic all-atom force field. *J. Phys. Chem. B* **108**, 2038–2047 (2004).
56. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
57. Patriksson, A. & van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **10**, 2073–2077 (2008).
58. Nadler, W. & Hansmann, U. H. E. Optimized explicit-solvent replica exchange molecular dynamics from scratch. *J. Phys. Chem. B* **112**, 10386–10387 (2008).
59. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
60. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

## Acknowledgements

We thank the National Natural Science Foundation of China (Grant No. 21373258), National Basic Research Program (Grant No. 2014CB910400, 2012CB910403) and National 863 Program (Grant No. 2012AA01A305) for financial support. The simulations were run at Shanghai Supercomputing Center (SSC) and TianHe 1 supercomputer in Tianjian.

### Author Contributions

Q.S., J.S. and W.Z. conceived the study. Y.Y. and J.W. performed the output analysis. Q.S. wrote the initial draft of the paper. All authors contributed to interpreting results and improvement of the paper.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Yu, Y. *et al.* The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: a microscopic view. *Sci. Rep.* **6**, 19500; doi: 10.1038/srep19500 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>