NIH Public Access
Author Manuscript
Proteins. Author manuscript; available in PMC 2012 June 1.

Published in final edited form as:
Proteins. 2011 June ; 79(6): 1704–1714. doi:10.1002/prot.22993.

The Ensemble Folding Kinetics of the FBP28 WW Domain Revealed by an All-atom Monte Carlo Simulation in a Knowledge-based Potential

Jiabin Xu[1], **Lei Huang**[1], and **Eugene I. Shakhnovich**[1]

[1] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge MA, 02138, USA

Abstract

In this work, we apply a detailed all-atom model with a transferable knowledge-based potential to study the folding kinetics of Formin-Binding protein, FBP28, which is a canonical three-stranded β-sheet WW domain. Replica exchange Monte Carlo (REMC) simulations starting from random coils find native-like (C α RMSD of 2.68Å) lowest energy structure. We also study the folding kinetics of FBP28 WW domain by performing a large number of *ab initio* Monte Carlo folding simulations. Using these trajectories, we examine the order of formation of two β –hairpins, the folding mechanism of each individual β– hairpin, and transition state ensemble (TSE) of FBP28 WW domain and compare our results with experimental data and previous computational studies. To obtain detailed structural information on the folding dynamics viewed as an ensemble process, we perform a clustering analysis procedure based on graph theory. Further, a rigorous P_{fold} analysis is used to obtain representative samples of the TSEs showing good quantitative agreement between experimental and simulated Φ values. Our analysis shows that the turn structure between first and second β strands is a partially stable structural motif that gets formed before entering the TSE in FBP28 WW domain and there exist two major pathways for the folding of FBP28 WW domain, which differ in the order and mechanism of hairpin formation.

Keywords

transition state ensemble; protein folding; β-strand; β-hairpin; β-sheet; Φ-value analysis; P_{fold} analysis

Introduction

Understanding the folding mechanism of β– structure is crucial for general and comprehensive understanding of protein folding kinetics. Compared to α-helical proteins, structure prediction and study of folding kinetics of β-proteins is more computationally challenging because β hairpin is an extended structure with a large number of long-range contacts, making it more difficult to reach its correct structure in an atomistic computer simulation.[1] Therefore, most simulation studies on folding of β-proteins are limited to small β-sheet domain, for example, the WW domain.[2–8] Formin-binding protein 28 WW domain (FBP28) is one member of the WW domain family. FBP28 is a small three-stranded β-sheet protein with high content of hydrophobic and aromatic residues. The characteristic features of WW domain are that this family of proteins has two highly conserved tryptophan residues and a strictly conserved proline residue. The native structure of FBP28 has been resolved by

*To whom correspondence should be addressed. eugene@belok.harvard.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NMR.[9–10] The FBP28 makes interactions with many signaling and regulatory proteins,[11] and can also form complexes which have been implicated in a number of diseases such as Alzheimer's and Huntington's disease.[12] FBP28 unfolds reversibly in both denaturant and thermal denaturation experiments[10,13–17], but it can also form amyloids at elevated temperature.[18] Temperature jump experiment showed that folding of FBP 28 is a cooperative, two-state process without any intermediate state detected.[13] Another laser-temperature jump experiment suggests that there are two decay phases for wild-type FBP28, the fast one is about 30μs and the slow one is >900μs at low temperature.[19] The heterogeneity suggests that a third state has to be considered in the folding process. Moreover, a large number of Φ values have been obtained experimentally by mutational analysis on FBP28, which may serve as a benchmark for simulation studies.[8]

Many computational studies have been performed on FBP28 or other family members of WW domain to gain insight into the formation of β structures. These studies can be grouped into the following categories: the first type of simulations employ high temperature unfolding [8] The drawback of this approach is that the reconstructed high-T folding pathways do not necessarily dynamically coincide with ones at ambient temperature [20–21]. The second type of simulations used replica exchange method, e.g. REMD (Replica Exchange Molecular Dynamics)[22] and multiplexed Q–replica molecular dynamics[3], to study equilibrium thermodynamics of the protein and derive the folding pathway indirectly from the free energy landscape. However the issue of how to derive dynamics from low-dimensiuonal projections of energy landscape remains unresolved [23–24]. The third type of simulations used the structure-based G model to directly study the folding dynamics at a fixed temperature from extended random coils[2,5]. There are no attractive non-native interactions in the G model which may be unphysical – several studies showed the importance of transient stabilizing non-native interactions at various stages of folding[25–27] Recent simulation used "physics-based" force field to study folding dynamics of WW domains at fixed temperature.[6,28] However, this method, while highly desirable, is still too computationally costly to produce sufficient number of folding events for detailed statistical anlysis.

Recently, we developed an all-atom knowledge-based potential, which succeeded in folding a diverse set of proteins to their near-native conformations.[29] In addition, our potential, combined with dynamic Metropolis Monte Carlo (MC) simulation methods, has been used to study folding dynamics of α-helical proteins directly from extended random coils at a fixed temperature.[30–31] Our group used structural kinetics cluster analysis in combination with transition state ensemble analysis and Φ value calculation to analyze folding pathways of α-helical proteins.[30,32] Good agreement with experiment suggests that this approach can reproduce folding dynamics of proteins efficiently and with good accuracy. The key feature of our approach is that it uses an all-atom model to provide an atomistically resolved picture of the folding process. However, it is somewhat coarse-grained dynamically making it efficient enough to generate a large number of long-time trajectories to glean statistically significant robust features of the folding process. Here we apply this approach to get insights into folding mechanism(s) of β– proteins using FBP28 as our model. There are several fundamental questions concerning folding of FBP28 as a prototypical β– protein. For instance, in what order are two β –hairpins formed in FBP28? What's the folding mechanism(s) of individual β hairpins? Are they the same or different? What's the TSE (transition state ensemble) and nucleation center during the folding process? The purpose of this paper is to address these questions by direct all-atom folding simulation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Models and Methods

The detailed description of the simulation model could be found elsewhere[29–30]. Here we give a brief summary of the model and simulation technique. First, all heavy-atom positions of the FBP28 WW domain were acquired from the NMR structure (residues 6–32 of Protein Data Bank id 1e0l), with the unstructured tails truncated.[10] The N-terminal sequence is conformationally flexible and does not interact with the remainder of the WW domain so the truncation of the N-terminal residues had no observable effect on the stability of the domain. [19] The truncation of the C-terminal residues decreases the stability of the protein because of the deletion of Leu-36, which forms a hydrophobic core with Trp-8, Tyr-20 and Pro-33 in the wild-type native state. Nevertheless, previous experimental work showed that the truncation does not result in significant structural change of the native state.[19] In addition, Periole et. al. performed simulation of the full-length FBP28 WW domain using three types of models(all-atom and explicit solvent, all-atom and implicit solvent, and C α-atom) and confirmed that the truncation does not affect the stability of the peptide.[33] Therefore, many simulation papers used the truncated version of the FBP28 WW domain, making simulation study more computationally accessible.[7,33] There are 27 residues and 238 atoms in total. In our model, Tyr-11, Tyr-19, Tyr-21 and Trp-30 form main hydrophobic core. Trp-8 and Tyr-20 form another hydrophobic core. The all-atom "knowledge-based" transferable energy function takes the form as:

$$E=w_{con}\times E_{con}+w_{trp}\times E_{trp}+w_{hb}\times E_{hb}+w_{sct}\times E_{sct} \quad [1]$$

where E_{con} is the pairwise atom-atom contact potential, E_{hb} is the hydrogen-bonding potential, E_{trp} is the sequence-dependent local torsional potential based on the statistics of sequential amino acid triplets, and E_{sct} is the side-chain torsional angle potential.

To test the ability of the potential to identify near-native state as lowest energy one, we use the REMC simulation to sample the conformation space with 32 replicas at different temperatures, ranging from 0.15 to 1.50. In the REMC simulation, we can move ψ and χ angles of all residues and φ except in proline and we use three different move sets to increase the sampling efficiency: backbone moves, side-chain moves, and "knowledge-based" moves. The backbone move has two types with equal probability: global move and local move. A global move is to rotate the dihedral angle (φ or ψ) of a randomly selected residue. A local move moves seven successive torsional angles with other residues unchanged. The step sizes of the global and local moves for the backbone are drawn from a normal distribution with zero mean and standard deviation of 2° and60°, respectively. A side-chain move consists of rotating all χ angles in a randomly selected nonproline residue. The step size of the side-chain rotation is drawn from a normal distribution with zero mean and standard deviation of 10°. The knowledge-based moves were discussed in details elsewhere.[34] A knowledge-based move of a residue during simulation entails setting the dihedral angles of the residue randomly to one of the clustered φ/ψ angles. The knowledge-based move can efficiently sample low energy states. For folding kinetics study, we perform 2304 independent Monte Carlo simulations, starting from different random coil configurations at T = 0.50 for 10^8 steps. The ensemble of initial random coil conformations is obtained by first running 5×10^5 MC steps at very high temperature, T = 1000 for each trajectory. Snapshots were stored at every 5×10^5 MC steps. Backbone moves and side-chains moves are still used in folding kinetics simulation. To satisfy the detailed balance condition, a knowledge-based move used in REMC simulation was not used, and the local move set was modified.[35] A new sampling method rather than the conventional Metropolis rule is used to conserve detailed balance. The probability of accepting a move from the old state o to the new state n for the local move set is given by

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$P(o \rightarrow n) = \min\left[1, \frac{N^{(n)}\exp(-U(n)/T)J(n)}{N^{(o)}\exp(-U(o)/T)J(o)}\right],$$

where N is the number of solutions, U is the potential energy, T is temperature, and J is the Jacobian determinant.

Not all of the 2304 trajectories contain native-like low-energy structures. Therefore, before turning to the folding kinetics, we make an initial objective selection of a set of "representative" trajectories. There is one minimum energy structure in each of the 2304 trajectories and we select 100 trajectories whose minimum energy structures have the lowest energies. To better quantify the structure similarity between the simulation structure and native structure, we use fraction of nonlocal native contacts (|i−j|>2) as our order parameter to monitor the folding process. Two residues are in contact if any two of their heavy atoms are in contact. Two heavy atoms are defined to be in contact if the distance between them is less than $\lambda(r_A + r_B)$, where r_A and r_B are their van der Waals radii and λ = 1.8.[29]

A simulated Φ value is defined according to Vendruscolo and co-workers [36] as

$$\Phi_I^{sim} = \frac{N_I^{TS}}{N_I^{NS}}$$

where N_I^{TS} is the average number of native contacts made by residue I in the transition state ensemble, and N_I^{NS} is the number of native contacts made by residue I in the native state.

Results

First, we check whether our potential can identify a set of near-native conformations of FBP 28 as global energy minimum. To that end, we performed replica exchange Monte Carlo (REMC) simulation with our energy function, starting from random coils. We obtained a total of 14719 structures and the energy landscape is shown in Figure 1(A). The minimum energy structure (Figure 1(B)) has the correct topology with three β strands correctly folded and a C α RMSD of 2.68Å. Some differences between the simulated lowest energy structure and the experimental structure are: first, Ser-6 has no contacts with other residues in the experimental structure, while it has contacts with Asn-23 and Arg-24 in the simulated structure. Second, Trp-8 has several contacts with Glu-27, Ser-28 and Thr-29 in the experimental structure, while such contacts are not observed in the simulated structure. Third, the β strand 3 in our simulated minimum energy structure is longer than that in the native structure. Importantly, our simulation correctly predicts two hydrophobic cores and side-chains belonging to these two hydrophobic cores are in the correct position. The results show the power of our knowledge-based potential to discriminate between near-native conformations and misfolded ones.

Folding Dynamics and Secondary Structure Formation

We selected 100 trajectories out of total 2304 for detailed analysis of folding dynamics. The temperature used in our dynamic Monte Carlo simulation is 0.5 in arbitrary units of temperature used in our simulations. We relate our temperature units to real temperature using the simulated melting curve simulation (Figure 2), which shows mid-transition at ~

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

0.6, while the experimental folding temperature of FBP28 is 337K.[8] Therefore, our simulation temperature of 0.5 corresponds to real temperature of ~281K.

The average fraction of total native contacts Q and the native contacts between β1 and β2, betweenβ2 and β3, within loop 1 and between loop 1 and other residues, and within loop 2 and between loop 2 and other residues (averaged over all of 100 folding trajectories) are shown as a function MC time-steps in Figure 3(A). The formation of the native contacts between β1 and β2 is faster than the formation of the native contacts between β2 and β3. Also, there is a rapid formation of these structures at early stages of folding.

To further understand the details of the folding process, we plot the probabilities of contacts at different stages of folding (Q between 0.0 and 0.5) in Figure 3(B–F). All 100 trajectories are used to make the plot in Figure 3. At $0.0<Q<0.1$, the contact pair between the Tyr-21 and Arg-24 has the highest contact probability (0.225). The majority of the contacts are neighboring contacts, indicating that the structures are still in the random coil state. At $0.1<Q<0.2$, the highest contact probability locates at loop1 region for β hairpin 1 and the contact probability decreases outward from the turn to the end of the hairpin. The highest contact probability for β hairpin 2 locates at loop2 region for β hairpin 2. At $0.2<Q<0.3$, the contact probability for β hairpin 1 continues to increase over 0.40 and the contact probability for β hairpin 2 has little change compared to $0.1<Q<0.2$, indicating that the formation of β hairpin 1 could occur earlier than the formation of β hairpin 2. At $0.3<Q<0.4$, the contact probability in the loop1 region increases to over 0.50 along with increased probabilities of other contacts within loop1 and between β1 and β2. For β hairpin 2, the contact probability increases to about 0.4 for two regions and in between these two green regions there is a blue region with a lower contact probability. At $0.4<Q<0.5$, the contact probabilities for pair residues in β hairpin 1 reach over 0.7 and the contact probabilities for pair of residues in β hairpin 2 are over 0.3. The above results suggest that statistically there are more folding pathways whereby the β hairpin 1 forms first and β hairpin 2 forms later. For the folding mechanism for each β hairpin, we observed that the contacts are first formed near the turn and then propagate outward for β hairpin 1. For β hairpin 2, the contacts first formed at two separate regions in the hairpin and later the whole β hairpin is formed.

Folding Mechanism for Individual β Hairpin Formation

It is worth noticing that a contact between two residues does not necessarily imply that there is a hydrogen bond between them. In order to see the formation of hydrogen bonds in both hairpins, we monitor eleven main chain H-bond contacts at different folding stages (Q values) shown in Figure 4(A) and Table 1. Since we use a heavy atom model, we measure the distance between the N atom and O atom of two residues to determine formation of a hydrogen bond. If the distance between the N atom and O atom is smaller than 3.5 Å, then we define that there is a hydrogen bond between these two residues. From Figure 4(B), we can see that the probability of H1 in β hairpin 1 is always highest from $0.0<Q<0.4$ and the probability decreases outward from the turn region to the end of the β hairpin 1, indicating that the formation of hydrogen bonds starts from the turn region to the end of the hairpin for β hairpin 1. For β hairpin 2, the probability is different, where the probability is lowest near the turn and it increases outward from the turn region to the end of the hairpin, indicating that the formation of hydrogen bond starts from the end of the hairpin to the turn region for β hairpin 2.

Structural Kinetic Cluster Analysis

In order to identify possible obligatory intermediates during the folding process, we use a structural cluster procedure developed before[32]. The cluster procedure uses a “structural graph” of geometrically clustered conformations to provide a coarse-grained structural and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

kinetic information during folding process. The structural clustering procedure is different from kinetic clustering employed by several authors[37–39] and is carried out in two steps. In the first step, all snapshots from 100 representative trajectories are clustered in a single-link graph. Each node in this graph represents a conformation. Two nodes are linked together by an edge if their structural similarity distance measure d is smaller than the cutoff value d_c. Therefore, we will get several clusters in our "structural graph" after the first step. The largest cluster, which contains near-native conformations, is called the Giant Component (GC). In the second step, an important quantity flux, F, which is defined as the fraction of all trajectories passing through the cluster, is introduced to characterize the clusters kinetically. Therefore, the clusters with high F constitute major folding intermediates (on or off-pathway). Clusters with F=1 are the set of conformations constituting obligatory intermediate states. In addition, we also calculate the mean-first passage time (MFPT) and the mean least-exit time (MLET) for each cluster. Finally, one representative structure, defined as the structure with the highest number of edges, is extracted from each cluster. These quantities, together with the representative structure from each cluster, provide a detailed picture of folding process from an ensemble perspective.

In this paper, we follow Hubner et al.[32] and use rmsd, distance rmsd(drms) and Rg as our order parameters for clustering. Each order parameter provides different complementary perspectives on the folding process. Table 1–3 in Supplemental Information show the results of structural kinetic cluster analysis for FBP 28 WW domain. When we use drms and rmsd as order parameters, we find only one single dominating cluster with high flux, which is the native state cluster. The absence of high flux clusters at early time of the folding process in the drms and rmsd structural cluster result means that at the initial stage of folding, there is no accumulation of a structurally well-defined folding intermediate. When we use Rg as order parameter, we observe a large number of clusters at early time of folding process with large variation of Rg. The largest cluster (GC) is a low-Rg cluster with MFPT $\approx 4\times10^6$ MC steps. However, the GC in the Rg cluster must contain not only conformations that are part of the native conformational ensemble but also pre-TSE low Rg conformations. We observe some representative structures with folded β –hairpin 2, but fragments of β1 form a small α–helix.(Figure 5) This type of structures is observed in a recent unfolding simulation using explicit solvent[7] and high temperature unfolding MD simulation.[8]

Transition State Ensembles

Transition state ensembles (TSEs) are key to understand the folding pathways. We use P_{fold} analysis to construct the TSEs from putative TSEs.[23] The P_{fold} analysis is based on the fact that simulations starting from a transition state conformation have equal probability of reaching the native state and a conformation belonging to the unfolded state. The way we get the putative TSEs is to select structures that immediately precede entry into the Giant Component (GC), which is the largest cluster in the RMSD structural cluster graph, which gives us 239 putative transition-state structures. For each conformation in the putative TSEs, we perform 256 independent short MC simulations with 10^6 MC steps. If the trajectory contains at least one structure whose RMSD to the minimum energy structure obtained from the REMC simulation is smaller than 3.5 Å, then we count this trajectory as a trajectory that reached the native state ensemble. Conformations with $0.4 < P_{fold} < 0.6$ constitute the TSE. This procedure generates a set of 15 "true" transition state structures for FBP 28 WW domain. (Figure 6)

There are 10 transition state structures having formed hairpin 1 of β1 and β2 with an unformed hairpin 2 of β2 and β3. Two transition state structures have a well-formed hairpin 2 of β2 and β3 with an unformed hairpin 1 of β1 and β2. The remaining 3 transition state structures do not have secondary structures formed. This type of transition state structures with no secondary structures formed are also observed in the previous study of high

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

temperature unfolding MD simulations.[8] The structural analysis of the transition state ensemble demonstrates that the dominant folding pathway is that first β hairpin forms first and second β –hairpin forms later. The minor folding pathway is that second β hairpin forms first and first β hairpin forms later.

Having obtained the true TSEs by P_{fold} analysis, we are now ready to use these structures to calculate thetheoretical Φ values for FBP 28 WW domain. Following previous conventions[36], Φ_i for a residue i is interpreted as the number of contacts present in the TSE for residue i divided by the number of native contacts(of the same residue i). Simulation Φ values with their standard deviations, averaged over all TSE conformations are given in Figure 7. Experimental Φ values have been obtained previously for FBP 28 WW domain.[8] The agreement between theory and experiment is good. Exceptions are Trp-8, Thr-9, Glu-10 and Ser-28 in our protein model, where the simulated Φ values are much higher than the experimental Φ values. The reason for the discrepancy is that there are very few native contacts for these residues in native structures so the simulated Φ values are not reliable – they can be very high and have large standard deviation. Another important reason is that our model uses implicit solvent. In reality these residues will form hydrogen bonds with water molecules while in simulation they will form other intramolecular contacts, resulting in apparently high Φ values. We find that, the most structured regions in the TSE is the turn between β1 and β2, as indicated by high Φ values, which forms a native-like β hairpin turn. There is another peak of Φ values in the region between β2 and β3, which suggests that the hairpin structure between β2 and β3 is also weakly formed. This picture is in good agreement with the result obtained by detailed all-atom high temperature unfolding molecular dynamics simulation.[8]

Discussion

Most probable folding pathways

Using a relatively simple transferable knowledge-based all-atom model, we performed a large number of *ab initio* protein folding runs for FBP28 WW domain that provided us with necessary data to study the folding kinetics as an ensemble process. By combining our results, we obtained a detailed picture of the folding dynamics of the three β-strand FBP28 WW domain. The dominant folding pathway includes first formation of β –hairpin 1 which consists of β1 and β2, followed by formation of β– hairpin 2 which consists of β2 and β3. The other non-dominant folding pathway is formation of β hairpin 2 first, followed by the formation of β hairpin 1. This non-dominant folding pathway was found earlier in improved Gō model simulations[40] and in a recent study using multiple rare event simulations.[7] Our finding of the propensity of hairpin 1 to form first during folding for FBP28 WW domain agrees with the result of Juraszek et al,[7] who found that the free energy barrier between unfolded states and intermediate state with only hairpin 1 formed is much lower than the free energy barrier between unfolded states and intermediate state with only hairpin 2 formed. In addition, our simulations qualitatively agree with the results by Luo et al. on Pin1 WW domain using the Gō model Molecular Dynamics simulation, which showed that Pin1 WW domain also has two folding pathways that differ by sequence in which hairpins are formed.[5] Our findings are also consistent with the simulation results by Ensign and Pande on the Fip35 in implicit solvent, in which it was found that the mechanism is heterogeneous, but that the larger hairpin (first) is more likely to form first.[41] Previous high-temperature unfolding simulation has shown that the contacts of the first β hairpin forming early in the folding process is the dominant folding pathway[8] and our result showed that this dominant pathway is still the same at ambient condition. Moreover, we also observed a structure with α-helix in the N-terminus with a relatively large Rg, which has been reported before in high temperature unfolding simulation[8] and bias-exchange metadynamics unfolding simulation[7]. A possible reason for the observation that dominant folding pathway involves an early

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

formation of hairpin 1 is that hairpin 1 has more aromatic residues, which belong to the hydrophobic core of the native protein, than hairpin 2. There are five aromatic hydrophobic residues (Trp-8, Tyr-11, Tyr-19, Tyr-20 and Tyr-21) involved in stabilizing β hairpin 1, while there are only two aromatic hydrophobic residues (Tyr-19 and Trp-30) involved in stabilizing β hairpin 2. Therefore, the assembly of β hairpin 1 is enthalpically more favorable than β hairpin 2. In addition, the lengths for loop 1 and loop 2 are almost the same so the entropic contributions are almost the same for two loops. Taken together these factors indicate that, - it is more likely that β hairpin 1 will get formed first.

Folding mechanism of two β hairpins

There are two proposed mechanisms for β hairpin folding. The first mechanism is the "zipper" model proposed by Munoz et al,[42] which involves the initial folding of the turn structure and following formation of hydrogen bonds zipping from the turn to the end of the hairpin. The other mechanism is the hydrophobic collapse mechanism proposed by Dinner et al, stating that the hydrophobic collapse nucleates the hairpin formation.[43] Our simulations show that the folding mechanism for β hairpin 1 follows the "zipper" model while the folding mechanism for β hairpin 2 follows the hydrophobic collapse mechanism. Previous study using high temperature unfolding method showed that the folding mechanism for β–hairpin 1 was hydrophobic collapse[8]. There are several possible explanations for the discrepancy between our results and the previous results. First, the previous simulation study was performed at high temperature (373K) and the native contacts for the hydrophobic interactions are more stable to withstand the thermal fluctuations than the native contacts at the turn area for β –hairpin 1. Therefore, previous high temperature unfolding simulation probably favored the hydrophobic collapse mechanism. Our simulation is performed at low temperature (~281K) and the first formation of hydrogen bonds near the turn is entropically more favorable because these contacts are spatially closer. It is therefore possible that the folding mechanism of β hairpin 1 is temperature dependent. At high temperature, the folding mechanism for β hairpin 1 may involve hydrophobic collapse and at low temperature, the folding mechanism for β hairpin 1 may follow the "zipper" model. Luo et al. used G model to fold Pin1 WW domain and found that β1 – β2 hairpin folded via a turn zipper mechanism at low temperatures but a hydrophobic collapse mechanism at the folding-transition temperature.[2] The difference of the folding mechanism for the two hairpins can be understood as follows. For β1 – β2 hairpin, the closest hydrogen bond to the turn region is between Thr-13 and Gly-16, which are only two residues apart. Therefore, it is relatively easy to get this hydrogen bond formed first due to spatial proximity. For β2– β3 hairpin, the closet hydrogen bond to the turn region is between Asn-22 and Glu-27, which are four residues apart. Therefore, it is relatively hard for this hydrogen bond to form first at low temperature. In this case, the hydrophobic interaction is the major driving force to form β2–β3 hairpin.

Transition State Ensemble and Nucleation Center

Our simulation suggests that the β-turn structure in the first β hairpin is the most structured region according to the result of Φ value analysis. We also observed relatively high Φ values in the β-turn region in the second β hairpin, which corresponds to transition state conformations with only β2 and β3 formed. Our prediction from simulated Φ values analysis agrees with the previous REMD simulation by Mu et al,[22] who predict the turn-1 formation as the transition state. However, we also get other types of "non-dominant" transition state structures in our simulation, e.g. transition states with no β structures formed, which were also observed in high temperature unfolding MD simulation.[8]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Conclusion

We use our transferrable knowledge-based energy potential to perform multiple folding trajectories, which allows us to get a complete picture of the folding kinetics from an ensemble perspective. Further, we use the most reliable method, the p_{fold} analysis, to identify the transition state ensembles and calculate simulated Φ values for all residues of the FBP28. The statistically significant number of folding events, combined with the structural cluster analysis technique, provides a complete and detailed outline of the ensemble pathway of the FBP28 WW domain folding, and possibly an insight into general features of kinetics of β-sheet formation. The conclusion we get from this study is that, first, there are two folding pathways for FBP28 WW domain. The dominant folding pathway involves formation of β hairpin 1 first, followed by the formation of β hairpin 2. The other non-dominant folding pathway is the first formation of β– hairpin 2, followed by the formation of β– hairpin 1; second, at low temperature, the folding mechanism for the two β–hairpins are different. β hairpin 1 follows the "zipper" folding mechanism and β– hairpin 2 follows hydrophobic collapse folding mechanism. Third, Φ-value analysis suggests that the turn region in β –hairpin 1 is the nucleation center and the transition state ensembles can be categorized as three types of conformations: 1) structures with β1 and β2; 2) structures with β2 and β3; 3) structures without secondary structures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jae Shick Yang, Peter Kutchukian, Adrian Serohijos and Lee Wei Yang for helpful discussions and comments on the manuscript. This work is supported by the NIH grant RO1 GM52126.

References

1. Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. Curr Opin Struct Biol. 2004; 14:76–88. [PubMed: 15102453]
2. Luo Z, Ding J, Zhou Y. Folding mechanisms of individual β–hairpins in a Gō model of Pin1 WW domain by all-atom molecular dynamics simulations. The Journal Of Chemical Physics. 2008; 128:225103–225110. [PubMed: 18554060]
3. Kim E, Jang S, Lim M, Pak Y. Free Energy Landscape of the FBP28 WW Domain by All-Atom Direct Folding Simulation. J Phys Chem B. 2010; 114:7686–7691. [PubMed: 20465282]
4. Sharpe T, Jonsson AL, Rutherford TJ, Daggett V, Fersht AR. The role of the turn in β-hairpin formation during WW domain folding. Protein Science. 2007; 16:2233–2239. [PubMed: 17766370]
5. Luo Z, Ding J, Zhou Y. Temperature-Dependent Folding Pathways of Pin1 WW Domain: An All-Atom Molecular Dynamics Simulation of a Go Model. Biophysical Journal. 2007; 93:2152–2161. [PubMed: 17513360]
6. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. Biophysical Journal: Biophysical Letters. 2008; 94:L75–L77.
7. Juraszek J, Bolhuis PG. (Un)Folding Mechanisms of the FBP28 WW Domain in Explicit Solvent Revealed by Multiple Rare Event Simulation Methods. Biophysical Journal. 2010; 98:646–656. [PubMed: 20159161]
8. Petrovich M, Jonsson AL, Ferguson N, Daggett V, Fersht AR. Φ-Analysis at the Experimental Limits: Mechanism of β-Hairpin Formation. J Mol Biol. 2006; 360:865–881. [PubMed: 16784750]
9. Chan DC, Bedford MT, Leder P. Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. EMBO J. 1996; 15:1045–1054. [PubMed: 8605874]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10. Macias MJ, Gervais V, Civera C, Oschkinat H. Structural analysis of WW domains and design of a WW prototype. Nat Struct Biol. 2000; 7:375–379. [PubMed: 10802733]
11. Ton-Lo W, Dongzhou H, Mohsen S, Kaizan H, Sudol M. Structure and function of the WW domain. Prog Biophys Mol Biol. 1996; 65:113–132. [PubMed: 9029943]
12. Hu H, Columbus J, Zhang Y, Wu D, Lian L, Yang S, Goodwin J, Luczak C, Carter M, Chen L, James M, Davis R, Sudol M, Rodwell J, Herrero JJ. A map of WW domain family interactions. Proteomics. 2004; 4:643–655. [PubMed: 14997488]
13. Ferguson N, Johnson CM, Macias M, Oschkinat H, Fersht A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. Proc Natl Acad Sci USA. 2001; 98:13002–13007. [PubMed: 11687613]
14. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M. The folding mechanism of a β-sheet: the WW domain. J Mol Biol. 2001; 311:373–393. [PubMed: 11478867]
15. Crane JC, Koepf EK, Kelly JW, Gruebele M. Mapping the transition state of the WW domain β-sheet. J Mol Biol. 2000; 298:283–292. [PubMed: 10764597]
16. Ferguson N, Pires JR, Toepert F, Johnson CM, Pan YP, Volkmer-Engert R, Schneider-Mergener J, Daggett V, Oschkinat H, Fersht A. Using flexible loop mimetics to extend Φ-value analysis to secondary structure interactions. Proc Natl Acad Sci USA. 2001; 98:13008–13013. [PubMed: 11687614]
17. Sudol M, Hunter T. NeW wrinkles for an old domain. Cell. 2000; 103:1001–1004. [PubMed: 11163176]
18. Ferguson N, Berriman J, Petrovich M, Sharpe TD, TFJ, Fersht AR. Rapid amyloid fiber formation from the fast-folding WW domain FBP28. Proc Natl Acad Sci USA. 2003; 100:9814–9819. [PubMed: 12897238]
19. Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. Proc Natl Acad Sci USA. 2003; 100:3948–3953. [PubMed: 12651955]
20. Finkelstein AV. Can protein unfolding simulate protein folding? Protein Eng. 1997; 10(8):843–845. [PubMed: 9415434]
21. Dinner AR, Karplus M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. J Mol Biol. 1999; 292(2):403–419. [PubMed: 10493884]
22. Mu YG, Nordenskiold L, Tam JP. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. Biophys J. 2006; 90:3983–3992. [PubMed: 16533840]
23. Du R, Pande V, Grosberg A, Tanaka T, Shakhnovich EI. On the transition coordinate for protein folding. Journal of Chemical Physics. 1998; 108(1):334–350.
24. Rao F, Settanni G, Guarnera E, Caflisch A. Estimation of protein folding probability from equilibrium simulations. J Chem Phys. 2005; 122(18):184901. [PubMed: 15918759]
25. Li L, Mirny LA, Shakhnovich EI. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. Nat Struct Biol. 2000; 7(4):336–342. [PubMed: 10742180]
26. Clementi C, Plotkin SS. The effects of nonnative interactions on protein folding rates: theory and simulation. Protein Sci. 2004; 13(7):1750–1766. [PubMed: 15215519]
27. Bowman GR, Pande VS. Protein folded states are kinetic hubs. Proc Natl Acad Sci U S A. 107(24):10890–10895. [PubMed: 20534497]
28. Maisuradze GG, Liwo A, Scheraga HA. Principal Component Analysis for Protein Folding Dynamics. J Mol Biol. 2009; 385:312–329. [PubMed: 18952103]
29. Yang JS, Chen WW, Skolnick J, Shakhnovich EI. All-Atom Ab Initio Folding of a Diverse Set of Proteins. Structure (London). 2007; 15:53–63.
30. Yang JS, Wallin S, Shakhnovich EI. Universality and diversity of folding mechanics for three-helix bundle proteins. Proc Natl Acad Sci USA. 2008; 105:895–900. [PubMed: 18195374]
31. Kutchukian PS, Yang JS, Verdine GL, Shakhnovich EI. All-Atom Model for Stabilization of α–Helical Structure in Peptides by Hydrocarbon Staples. J Am Chem Soc. 2009; 131:4623–4627.
32. Hubner IA, Deeds EJ, Shakhnovich EI. Understanding ensemble protein folding at atomic detail. Proc Natl Acad Sci USA. 2006; 103:17747–17752. [PubMed: 17095606]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

33. Periole X, Allen LR, Tamiola K, Mark AE, Paci E. Probing the free energy landscape of the FBP28WW domain using multiple techniques. J Comput Chem. 2009; 30:1059–1068. [PubMed: 18942730]
34. Chen WW, Yang JS, Shakhnovich EI. A Knowledge-Based Move Set for Protein Folding. Proteins: Structure, Function, and Bioinformatics. 2007; 66:682–688.
35. Dodd LR, Boone TD, Theodorou DN. A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. Mol Phys. 1993; 78:961–996.
36. Paci E, Vendruscolo M, Dobson CM, Karplus M. Determination of a transition state at atomic resolution from protein engineering data. J Mol Biol. 2002; 324:151–163. [PubMed: 12421565]
37. Rao F, Caflisch A. The protein folding network. J Mol Biol. 2004; 342:299–306. [PubMed: 15313625]
38. Bowman GR, Huang X, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. Methods. 2009; 49:197–201. [PubMed: 19410002]
39. Karpen ME, Tobias DJ, Brooks CLr. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. Biochemistry. 1993; 32:412–420. [PubMed: 8422350]
40. Karanicolas J, Brooks CL. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J Mol Biol. 2003; 334:309–325. [PubMed: 14607121]
41. Ensign DL, Pande VS. The Fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations. Biophys J. 2009; 96:L53–L55. [PubMed: 19383445]
42. Munoz V, Henry ER, Hofrichter J, Eaton WA. A statistical mechanical model for β-hairpin kinetics. Proc Natl Acad Sci USA. 1998; 95:5872–5879. [PubMed: 9600886]
43. Dinner AR, Lazaridis T, Karplus M. Understanding beta-hairpin formation. Proc Natl Acad Sci USA. 1999; 96:9068–9073. [PubMed: 10430896]
44. DeLano, WL. The PYMOL Molecular Graphics System. DeLano; San Carlos, CA: 2002.

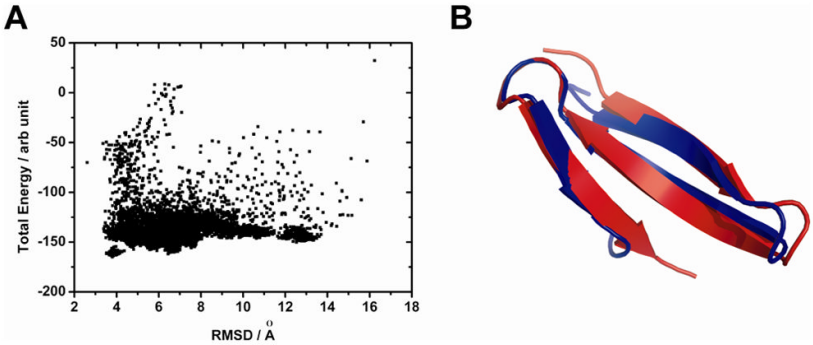

Fig 1.

Fig. 1(A) The energy landscape for FBP 28 WW domain in Ab initio REMC simulations as projected onto RMSD axis.

Fig. 1(B) Superposition of the backbones of the native structure (in blue) and minimum energy structure (in red) obtained through the REMC simulations. The RMSD and C α RMSD between the minimum energy structure and the native structure are 3.79 Å and 2.68Å, respectively. Structures were created by using PyMOL[44]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

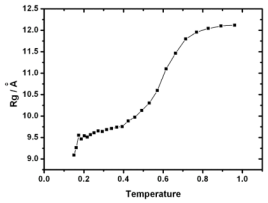

Fig. 2.
Simulated melting curve for the FBP28 WW domain in terms of average size of the molecule (Rg) vs temperature

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

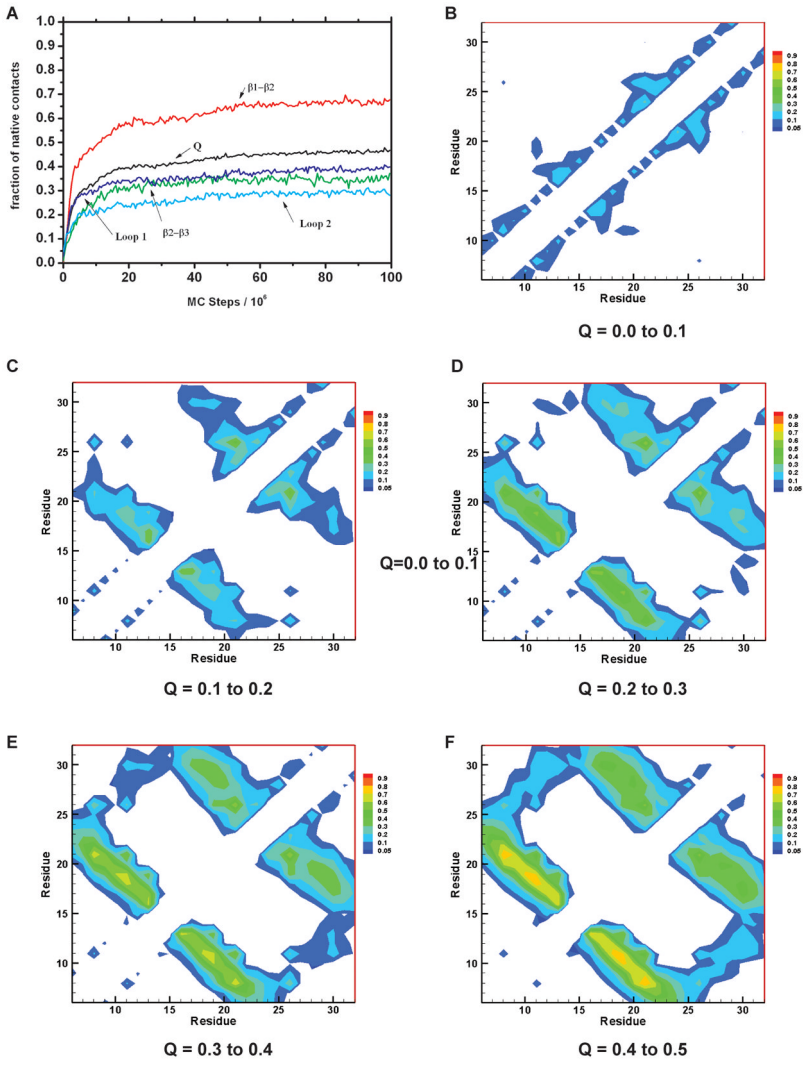

Fig 3.

Fig. 3(A) Fractions of native contacts averaged over all 100 trajectories as a function of MC time steps at T=0.50. The total fraction of native contacts (Q) is shown in black. The fraction of native contacts between β1 and β2 is in red, between β2 and β3 is in green, within loop 1 and between loop 1 and other residue is in blue, and within loop 2 and between loop 2 and other residues is in cyan.

Fig. 3(B–F) Probabilities of native residue-residue contact at various stages of folding according to the Q values. The folding temperature is 0.50.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

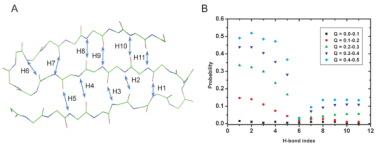

Fig 4.

Fig. 4(A) Eleven hydrogen bonds for β1 and β2 which are monitored during MC simulations.

Fig. 4(B) Probabilities of 11 H-bonds at various stages of folding categorized according to the Q values of 100 folding trajectories at T=0.50. The H-bond indices are defined in the text.

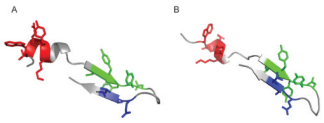

Fig. 5.
Two representative structures from two large Rg clusters with α-helical structures at N-terminus.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

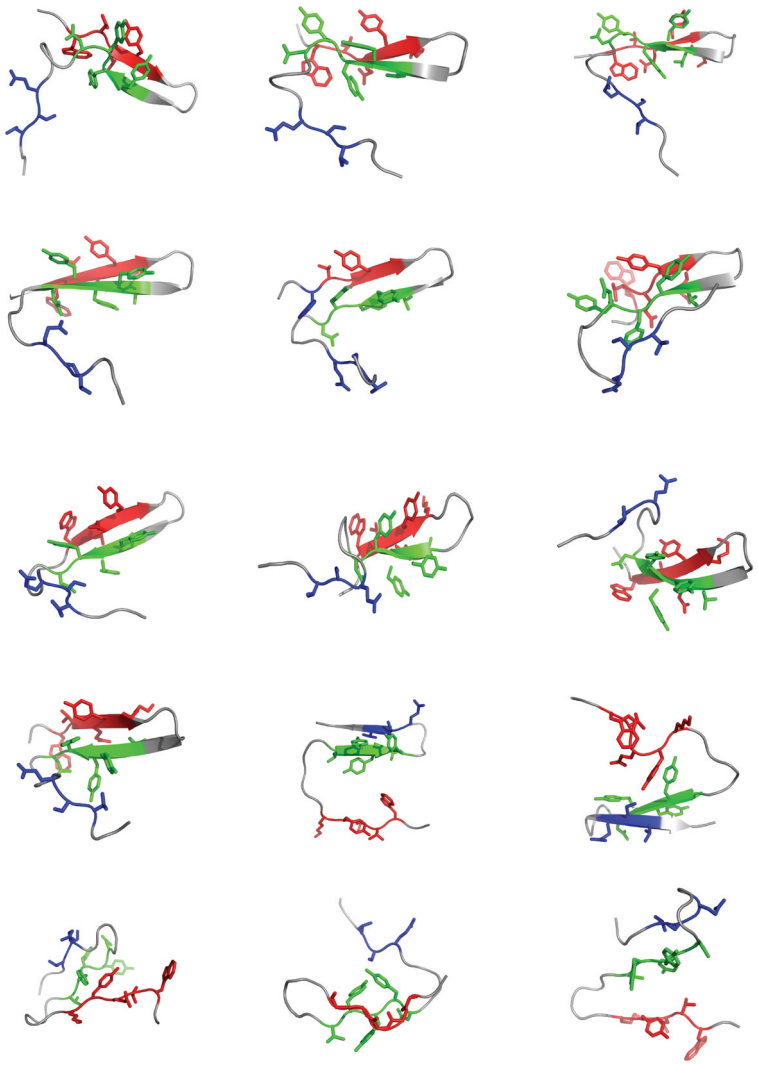

Fig. 6.
The transition state ensemble of 15 structures determined by the Pfold analysis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

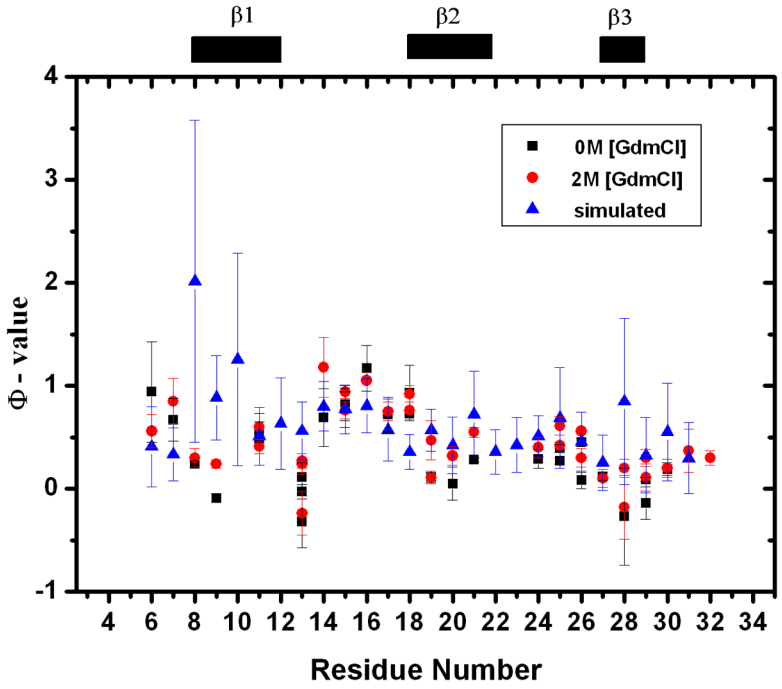

Fig. 7.

Comparison between simulated and experimental Φ values. Error bars denote the standard deviation σ of 15 Φ values calculated from 15 structures of transition state ensemble.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Eleven hydrogen bonds monitored during folding simulation

H1	Thr-13-N --- Lys-17-O
H2	Tyr-19-N --- Tyr-11-O
H3	Tyr-11-N --- Tyr-19-O
H4	Tyr-21-N --- Thr-9-O
H5	Thr-9-N --- Tyr-21-O
H6	Glu-27-N --- Asn-22-O
H7	Asn-22-N --- Glu-27-O
H8	Thr-29-N --- Tyr-20-O
H9	Tyr-20-N --- Thr-29-O
H10	Glu-31-N --- Thr-18-O
H11	Thr-18-N --- Glu-31-O