Folding mechanisms of individual $\beta$-hairpins in a Gø model of Pin1 WW domain by all-atom molecular dynamics simulations

Zhonglin Luo,1 Jiandong Ding,1,a and Yaoqi Zhou2,a
1Key Laboratory of Molecular Engineering of Polymers of Ministry of Education, Department of Macromolecular Science, Laboratory of Advanced Materials, Fudan University, Shanghai 200433, People’s Republic of China
2Indiana University School of Informatics and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Walker Plaza, 719 Indiana Av. Suite 319, Indianapolis, Indiana 46202, USA

(Received 21 March 2008; accepted 2 May 2008; published online 10 June 2008)

This paper examines the folding mechanism of an individual $\beta$-hairpin in the presence of other hairpins by using an off-lattice model of a small triple-stranded antiparallel $\beta$-sheet protein, Pin1 WW domain. The turn zipper model and the hydrophobic collapse model originally developed for a single $\beta$-hairpin in literature is confirmed to be useful in describing $\beta$-hairpins in model Pin1 WW domain. We find that the mechanism for folding a specific hairpin is independent of whether it folds first or second, but the formation process are significantly dependent on temperature. More specifically, $\beta1$-$\beta2$ hairpin folds via the turn zipper model at a low temperature and the hydrophobic collapse model at a high temperature, while the folding of $\beta2$-$\beta3$ hairpin follows the turn zipper model at both temperatures. The change in folding mechanisms is interpreted by the interplay between contact stability (enthalpy) and loop lengths (entropy), the effect of which is temperature dependent. © 2008 American Institute of Physics. [DOI: 10.1063/1.2936832]

I. INTRODUCTION

The folding mechanisms of $\beta$-hairpins have been known to be determined by the interplay between hydrophobic interactions and hydrogen-bonding interactions. One mechanism was proposed by Muñoz and co-workers based on the temperature-jump experiments and a statistical mechanical model of the GB1 hairpin, a 16-residue peptide. In this model, folding is initiated from the turn nucleation, followed by zipping the hairpin from the turn region to the end. This mechanism was later called the turn zipper mechanism. The second mechanism was postulated by Dinner et al. based on their multicanonical Monte Carlo simulations. Their simulations revealed that a hydrophobic collapse occurs prior to the hairpin formation, and it was called the hydrophobic collapse mechanism.

While the folding of a single $\beta$-hairpin has been studied in detail, it is not yet clear regarding how an individual $\beta$-hairpin would fold in the presence of other $\beta$-hairpins in the same protein. This paper focuses on the individual folding mechanism of each $\beta$-hairpin in a two-hairpin polypeptide by using Pin1 WW domain as the model. The WW domain family, named after two highly conserved tryptophans in the strands, is a triple-stranded antiparallel $\beta$-sheet protein with 30–50 residues. The WW domain family, which is stable in monomer and amenable to extensive mutagenesis, is regarded as one of the simplest $\beta$-sheet models for testing a variety of hypotheses regarding $\beta$-sheet folding. Pin1 WW domain, one of the best characterized peptide of this family, is the N-terminal (residues 1–39) of a two-domain protein hPin1, in which the C-terminal cis/trans-isomerase domain (residues 45–167) is weakly connected to the WW domain by a flexible linker. The truncation of residues 1–5 of Pin1 WW domain has no substantial influence on its function or stability. The structure of residues 6–39 is shown in Fig. 1. The overall topology of the Pin1 WW domain is made of three strands ($\beta1$, residues 11–15; $\beta2$, residues 22–26; and $\beta3$, residues 31–34) connected by two loops (loop 1, residues 16–21 and loop 2, residues 27–30). The protein could tolerate mutations almost expanding all strands and loops. The abnormal large $\phi$-values (>1) of loop 1 have been interpreted that the nucleation of loop 1 is the driving force to form $\beta1$-$\beta2$ hairpin, analogous to a turn zipper mechanism proposed by Muñoz and co-workers. Recent NMR measurements of Pin1 WW and other WW domains (hYAP56 and FBP28) indicated that loop 1 was dynamic at the folded state. Meanwhile, the unfolding simulation of FBP28 indicated that loop 1 was more like a hinge to bring the long-range hydrophobic residues together, although $\phi$-values of residues in loop 1 are larger than 1.

The folding kinetics of Pin1 WW domains has been studied theoretically on the molecular and atomic levels. With a three-letter code residue sequence (the hydrophobic, hydrophilic, and neutral), Cecconi et al. studied the thermodynamics of Pin1 WW domain. They reported that the folding and unfolding were essentially reversible. Its folding started from collapsed hydrophobic cores in the middle of $\beta1$-$\beta2$ hairpin and $\beta2$-$\beta3$ hairpin. Karanicolas et al., on the other hand, found that the $\beta2$-$\beta3$ hairpin of Pin1 WW domain folded prior to hairpin $\beta1$-$\beta2$ by using an improved $C_\alpha$-based model. Moreover, the difference in side-chain
The contact map of the global-minimum-energy structure of the model protein is represented as follows:

\[ u_{ij}^\text{bond}(r) = \begin{cases} 
\infty, & r < 0.9\sigma_{ij} \\
0, & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\
\infty, & r > 1.1\sigma_{ij},
\end{cases} \]

where \( \sigma_{ij} \) is the atomic distance obtained by the CHARMM energy minimization as mentioned above. The bond-length tolerance of 0.1 is introduced to decouple multibody collisions into binary collisions.

The square-well nonbonded \( i, j \) pair interactions are represented as follows:

\[ u_{ij}(r) = \begin{cases} 
\infty, & r < 0.8\sigma_{ij} \text{vdW} \\
B_{ij}e, & 0.8\sigma_{ij} \text{vdW} < r < 1.2\sigma_{ij} \text{vdW} \\
0, & r > 1.2\sigma_{ij} \text{vdW},
\end{cases} \]

where \( \sigma_{ij} \text{vdW} \) is the van der Waals distance obtained from the CHARMM polar hydrogen parameter set 19 (Ref. 46) and \( B_{ij} \) is the square-well depth. The factor of 0.8 is the typical ratio between the hard-core diameters and the van der Waals parameters.

While the relative sequential folding of the two hairpins in a WW domain has been studied in detail, little attention was paid to the detailed folding mechanisms for individual hairpins. Several fundamental questions are yet to be addressed. For instance, can the folding mechanism of a single \( \beta \) hairpin be applied to a \( \beta \)-hairpin within a protein (i.e., in the presence of other secondary-structure elements)? Can different \( \beta \) hairpins in a protein have different folding mechanisms? Is the folding mechanism of a \( \beta \) hairpin dependent on whether it is formed first or second? The purpose of this paper is to address these questions by folding and unfolding simulations of an all-atom Gō model of Pin 1 WW domain.

II. METHODS AND SIMULATIONS

A. Model

The initial heavy-atom positions of residues 6–39 of the native state were obtained from the x-ray structure coded 1pin.pdb in Protein Data Bank and the positions of the polar hydrogen atoms were generated by CHARMM program and minimized for 100 steps to remove bad contacts using the CHARMM polar hydrogen parameter set 19 with distance-dependent dielectric constants. The total number of all heavy atoms and polar hydrogen atoms is 361.

The interaction of two bonded atoms, as well as any 1,3 angle-constrained atomic pair and 1,4 aromatic carbon pair, \( i \) and \( j \), is described as an infinitely deep square-well potential

\[ u_{ij}^\text{bond}(r) = \begin{cases} 
\infty, & r < 0.9\sigma_{ij} \\
0, & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\
\infty, & r > 1.1\sigma_{ij},
\end{cases} \]

where \( \sigma_{ij} \) is the atomic distance obtained by the CHARMM energy minimization as mentioned above. The bond-length tolerance of 0.1 is introduced to decouple multibody collisions into binary collisions.

The square-well nonbonded \( i, j \) pair interactions are represented as follows:

\[ u_{ij}(r) = \begin{cases} 
\infty, & r < 0.8\sigma_{ij} \text{vdW} \\
B_{ij}e, & 0.8\sigma_{ij} \text{vdW} < r < 1.2\sigma_{ij} \text{vdW} \\
0, & r > 1.2\sigma_{ij} \text{vdW},
\end{cases} \]

where \( \sigma_{ij} \text{vdW} \) is the van der Waals distance obtained from the CHARMM polar hydrogen parameter set 19 (Ref. 46) and \( B_{ij} \) is the square-well depth. The factor of 0.8 is the typical ratio between the hard-core diameters and the van der Waals parameters.
rameters in the Weeks–Chandler–Andersen perturbation theory, while a ratio of 1.5 between the square-well and hard-core diameters is typical for small molecular systems.49

The L-form chirality about a tetrahedral extended heavy atom (\(\alpha_0=35.26439^\circ\)) as well as the plane of aromatic atoms (\(\alpha_0=0^\circ\)) are maintained by a “bondlike” potential for improper dihedral angels. A 20° fluctuation is used to increase the flexibility and folding rate

\[
\omega_{\text{improp}} = \begin{cases} \infty, & \omega - \omega_0 > 20^\circ \\ 0, & -20^\circ \leq \omega - \omega_0 \leq 20^\circ \\ \infty, & \omega - \omega_0 < -20^\circ \end{cases}
\] (3)

The CHARMM minimized structure generated as the above further undergoes a short discontinuous molecular dynamics (MD) simulation to remove all initial hard-core overlaps. The square-well interactions \(B_{ij}\) for the original native contacts found in the x-ray structure are set to be −100, which is a large negative energy in order to ensure that these interactions are maintained throughout the simulation. During the process, we modify the initially overlapped hard-core diameters found in the CHARMM minimized structure until the true hard-core diameters are reached. The resulting structure has a main-chain root-mean-squared deviation (RMSD) from the original x-ray structure of 0.69 Å and is regarded as the global minimum of the CHARMM minimized structure. A 20° fluctuation is used to investigate the folding mechanisms. Two residues are in contact if there is any square-well atomic overlap between them.

The approximate transition temperature \(T^*=3.6\), only limited folding and unfolding events were obtained after long equilibrium simulations, and these trajectories were also used to investigate the folding mechanisms. A progress variable \(Q\), the fraction of nonlocal native contacts (\(|i-j|\geq 3\)) between residues [total number is 65, shown in Fig. 1(e)], is used to monitor the folding process. Two residues are in contact if there is any square-well atomic overlap between them.

III. RESULTS

A. Refolding at a low temperature \(T^*=3.0\)

The folding process of the model protein was triggered by a sudden drop of temperature from the equilibrium simulation at \(T^*=4.0\) to \(T^*=3.0\). We found that 102 out of 135 independent runs folded within the cutoff time of \(t^*=200\,000\). Here, a trajectory is considered as “folded” if the main-chain RMSD values of more than 20 configurations saved at an interval of \(t^*=100\) during simulations are smaller than 2.5 Å.

The average fractions of native contacts \(Q\) and the native contacts between \(\beta1\) and \(\beta2\), between \(\beta2\) and \(\beta3\), 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 135 folding trajectories) are shown as a function of the reduced time in Fig. 2. There is an initial formation of the native contacts in loop 2 which appears to trigger the quick formation of the contacts between \(\beta2\) and \(\beta3\). The formation of the native contacts of loop 1 is more gradual and faster than that of the contacts between \(\beta1\) and \(\beta2\) for \(t^*<125\,000\). However, it is only an average picture.

A detailed inspection of individual folding trajectories indicated that in the 102 folded runs, 80 trajectories fold with \(\beta2-\beta3\) hairpin formed first, which we call path I, and 22 trajectories fold with \(\beta1-\beta2\) hairpin formed first, which we call path II. The typical snapshots in path I and path II are illustrated in Fig. 3. We also inspected the 33 unfolded runs (until \(t^*=200\,000\)) and found that those trajectories can also be classified either with a mostly formed \(\beta2-\beta3\) hairpin or
$\beta_1-\beta_2$ hairpin. This suggests that those trajectories will, if running for a sufficiently long time, fold in similar folding pathways as those folded.

To further understand which contacts trigger the folding, we plot the probabilities of contacts at different stages of folding ($Q$ between 0.0 and 0.6) for path I (in the upper triangle) and path II (in the lower one) in Fig. 4. All 102 folded trajectories in paths I and II are used to make the plot in Fig. 4. We did not show the result for $Q \geq 0.6$ because folding is essentially completed for $Q \geq 0.6$. We first take a look at the $\beta_1-\beta_2$ hairpin in path I. At $0.1 < Q < 0.2$, the contact pair between the S16 and S19 (loop 1) has the highest contact probability (0.63). This was followed by the contact probability of 0.22 between S16 and R21. At $0.2 < Q < 0.3$, the contact probability between S16 and S19 further increases to 0.85 along with increased probabilities of other contacts within loop 1 (S16-R21) and between $\beta_1$ and $\beta_2$ (M15-V22 and R14-Y23). At $0.3 < Q < 0.4$, the central contact between $\beta_1$ and $\beta_2$ (R14-Y23) becomes stabilized and the folding of the $\beta_1-\beta_2$ hairpin is mostly completed. On the other hand, although the contact probability of pair N26-T29 in the $\beta_2-\beta_3$ hairpin is already 0.54 for $0.0 < Q < 0.1$, it does not change much during the folding process from $Q \sim 0.1$ to $Q \sim 0.4$. Thus, in path I and at $T^\alpha = 3.0$, the S16-S19 contact triggers the formation of $\beta_1-\beta_2$ hairpin, while the N26-T29 contact initiates the formation of hairpin $\beta_2-\beta_3$ after the formation of $\beta_1-\beta_2$ hairpin. Both S16-S19 and N26-T29 are located at the turns of the two hairpins, as shown in Fig. 1(b). Hence, both hairpins in path I can be viewed as folded via the turn zipper mechanism.

Path II is the dominant pathway at $T^\alpha = 3.0$ and the contact maps at different stages of folding are showed in the lower triangle of Fig. 4. The formation of the $\beta_2-\beta_3$ hairpin is early essentially completed as $0.2 < Q < 0.3$. In contrast, the formation of the $\beta_1-\beta_2$ hairpin did not occur until $0.4 < Q < 0.6$. Although the $\beta_2-\beta_3$ hairpin folded much earlier than the $\beta_1-\beta_2$ hairpin in path II, the folding processes of individual hairpins in paths I and II are the same (triggered by the same key contacts).

Figure 5 further analyzes the folding processes according to ten main-chain H-bond contacts and one side-chain H-bond at different folding stages ($Q$ values). There are H1—H4 for the contacts between $\beta_1$ and $\beta_2$, H5 and H6 within loop 1, H8 within loop 2, H9—H11 between $\beta_2$ and $\beta_3$, and H7, the side-chain H-bond between the N—H of T29 and the O==CNH$_2$ group of N26 [Fig. 1(b)].
In path I, the probability of H6 in the β1-β2 hairpin is the highest from 0.0 to 0.2. This is followed by H5 at 0.2 to 0.3, and then H4 at 0.3 to 0.6. Similarly, the side-chain H7 in loop 2 of the β2-β3 hairpin has the highest contact probability in path I. The main-chain H8 of N30 to O=C of N26 at the end of loop 2 is always lower than 0.2 because N30 has a very weak interaction with all other residues. In path II, Fig. 5(b) also shows that H6 is the first formed H-bond in the β1-β2 hairpin, and H7 is the first formed one in the β2-β3 hairpin. This confirms the turn zipper mechanism for two hairpins in two different folding paths.

B. Refolding and unfolding around the critical transition temperature \( T^* = 3.6 \)

We further explored the folding and unfolding events at \( T^* = 3.6 \), the folding transition temperature, at which the energy exhibited a bimodal distribution, as seen in Fig. 6, indicating a cooperative two-state transition between a coil state and a folded state. These events are difficult to observe at the transition temperature because of the equal stabilities of the folded and unfolded states. We obtained eight folding and unfolding equilibrium trajectories after very long simulations (on average, 40 days of CPU 2.8G for one trajectory lasting at \( t^* = 4000000 \)). Only path I (β1-β2 hairpin folds first) and the reverse of path I (β1-β2 hairpin unfolds last) were observed in these trajectories.

The contact maps of the configurations collected from folding and unfolding events at the transition temperature are shown in Fig. 7. To increase the statistics, we included 500 configurations before and after each folding (or unfolding) event. We found that the folding processes at the transition temperature essentially follow the reverse processes of unfolding. This is expected because the equilibrium folding/
unfolding simulations should be reversible. Although S16-S19 in loop 1 of β1-β2 hairpin has the highest contact probability at \( 0.0 < Q < 0.1 \), it is the contact between W11 and N26 that emerges with the second highest contact probability at \( 0.1 < Q < 0.2 \) and has the highest probability for \( Q > 0.2 \). So, the W11-N26 contact serves as the nucleation site for the growth and final folding of β1-β2, unlike the low-temperature folding that starts from loop 1. On the other hand, the folding mechanism of β2-β3 hairpin is the same as that observed at low temperatures: N26-T29 is the first contact pair to trigger the formation of β2-β3. The difference of the folding mechanisms of β1-β2 hairpin at different temperatures is highlighted in Fig. 8.

IV. DISCUSSION

A. Folding mechanisms for individual hairpins

The folding pathways of Pin 1 WW domain can be classified into two different paths based on the early formation of either β1-β2 hairpin (path I) or β2-β3 hairpin (path II). Path II dominates at low temperatures, while only path I prevails at the transition temperature. Detailed examinations indicate that the folding mechanism of an individual β hairpin is independent of whether it folds first or second (i.e., in path I or path II). Instead, it is somewhat dependent on the temperature of the folding simulation. While β2-β3 hairpin folds via a turn zipper mechanism at low or high temperatures, β1-β2 hairpin folds via a turn zipper mechanism at low temperatures but a hydrophobic collapse mechanism at the folding-transition temperature.

The above change in the mechanism can be understood as follows. At high temperatures, a high kinetic energy allows a protein to explore more conformational space and only those native contacts that are sufficiently stable to withstand thermal fluctuation can serve as the nucleation sites for folding. There are 52 native atomic interactions for the residue pair W11-N26 for stabilizing β1-β2 hairpin, the strongest residue-residue interaction in the model Pin1 WW domain. By comparison, the strongest residue-residue interaction in β2-β3 hairpin (between Y24 and Q33) has only 26 atomic interactions. This explains the collapse mechanism for the first formation of β1-β2 at the relatively high temperature \( (T^* = 3.6) \). Once β1-β2 hairpin is formed, β2-β3 hairpin can be assembled with the least entropic cost via the turn zipper mechanism. At low temperatures, each native contact is less likely broken as soon as it is formed. As a result, the contacts between the nearest neighbors might form first and trigger the folding event. This explains the turn zipper mechanism for both hairpins at the low temperature \( (T^* = 3.0) \). It also explains the dominance of path II at the low temperature: β2-β3 hairpin has a shorter loop and its assembly is entropically more favorable than β1-β2 hairpin with a longer loop.

B. Comparison with experimental results

The above interpretation agrees with the experiments that reducing the length of loop 1 and substituting with high turn-propensity residues lead to fast folding and more stabilized proteins.27,28,36 The difficulty of folding loop 1 found in simulations is consistent with the lack of strong interactions of loop 1 residues with other residues27,36 and the functional role of a flexible loop 1 for ligand binding.36 The theoretical finding that V22 and M15 are not important in folding agrees with the weaker preference for Met and Val as cross-strand neighbors in β-sheets found by experiments and statistics.55,56 The proposed nucleating W11-N26 contact is found to be critical for stabilizing Pin 1 WW domain.27

Based on mutation experiments and \( \phi \)-value analysis, Gruebele and co-workers suggested two distinct kinds of residues in Pin1 WW domain.25,27,28 loop 1 is responsible for folding rate and hydrophobic clusters for thermodynamic stability. At high temperatures, the hydrophobic clusters come into play but do not fully dominate the folding rate.25,27 Our simulations at different temperatures are consistent with these findings: at low temperatures, path II dominates and the folding of β1-β2 hairpin is the rate determining step, while at high temperatures, the hydrophobic contact between...
N26 and W11 becomes the most important nucleation contact. In our simulations, the contacting pair N26-T29 plays an important role in initiating the folding of the loop 2 (Figs. 4, 5, and 7). This is also consistent with the experimental observation that N26 is the structural gatekeeper in the WW domain. This may explain, in part, why the residue N26 is highly conserved in WW domain family. 

It should be emphasized, however, that a Gō model has its limitation because it contains only the native contact interactions. For example, it unlikely provides a good description of unfolded conformations where non-native contacts often play an important role. The purpose of this paper is to show the behavior of individual hairpins of Pin1 WW domain if its native contact energies dominate over non-native energies. This assumption is supported by the above-described consistency with some experimental results. However, it may not hold up as more experimental data become available. Nevertheless, the interplay between contact stability (enthalpy) and loop lengths (entropy) likely plays an important role in folding of many proteins.

C. Comparison with other simulation results

By analyzing the thermodynamic free energy profile based on a Cα model, Karanicolos and Brooks III predicted that the β2-β3 hairpin of Pin1 WW domain folded prior to β1-β2 hairpin. We found this pathway as the dominant path in our low-temperature folding simulations. The unfolding all-atom simulations of FBP (Ref. 29) indicated that β2-β3 hairpin was dissociated first and the hydrophobic contacts in β1-β2 hairpin were more stable than those in loop 1. This is similar to the reverse of path I discovered in our simulation. Cecconi et al. 32 suggested that the first contacts formed during folding were located in the middle of the two hairpins. During the unfolding reaction, the contacts of loop 1 and loop 2 disappeared at last. We did not find a similar pathway in our simulations. This is likely caused by the difference in models (a Cα model used by Cecconi et al. versus an all-atom model used in this study).

V. CONCLUSIONS

This paper has confirmed that the present two models of β-hairpin formation, the turn zipper model proposed by Mofouz and co-workers and hydrophobic collapse model proposed by Dinner et al. in their studies of a single hairpin in a rather short polypeptide, are still justified in description of kinetic processes of individual hairpins in a two-hairpin protein. We also found that two β-hairpins in the Gō model of Pin1 WW domain obeyed different and temperature-dependent folding mechanisms: at a low temperature, both hairpins fold via the turn zipper model; however, it is the hydrophobic interaction at the end of β1-β2 hairpin that initiates the folding of β1-β2 hairpin at the folding-transition temperature (thus following the hydrophobic collapse model). The result highlights the interplay of the entropic and enthalpic effects on folding at different temperatures.