Published in final edited form as:

J Struct Biol. 2011 November ; 176(2): 143–150. doi:10.1016/j.jsb.2011.08.001.

The Effect of Context on the Folding of β -Hairpins

Amanda L. Jonsson and Valerie Daggett*

Department of Bioengineering, University of Washington, Box 355013, Seattle, WA 98195-5013, USA

Abstract

Small β -hairpin peptides have been widely used as models for the folding of β -sheets. But how applicable is the folding of such models to β -structure in larger proteins with conventional hydrophobic cores? Here we present multiple unfolding simulations of three such proteins that contain the WW domain double hairpin β -sheet motif: cold shock protein A (CspA), cold shock protein B (CspB) and glucose permease IIA domain. We compare the behavior of the free motif in solution and in the context of proteins of different size and architecture. Both Csp proteins lost contacts between the double-hairpin motif and the protein core as the first step of unfolding and proceeded to unfold with loss of the third β -strand, similar to the isolated WW domain. The Glucose permease IIA domain is a larger protein and the contacts between the motif and the core were not lost as quickly. Instead the unfolding pathway of glucose permease IIA followed a different pathway with $\beta 1$ pulling away from the sheet first. Interestingly, when the double hairpin motif was excised from the glucose permease IIA domain and simulated in isolation in water it unfolded by the same pathway as the WW domain, indicating that it is tertiary interactions with the protein that alter the motif's unfolding not a sequence dependent effect on its intrinsic unfolding behavior. With respect to the unfolding of the hairpins, there was no consistent order to the loss of hydrogen bonds between the β -strands in the hairpins in any of the systems. Our results show that while the folding behavior of the isolated WW domain is generally consistent with the double hairpin motif's behavior in the cold shock proteins, it is not the case for the glucose permease IIA domain. So, one must be cautious in extrapolating findings from model systems to larger more complicated proteins where tertiary interactions can overwhelm intrinsic behavior.

Keywords

protein folding; β -hairpin; molecular dynamics; WW domain; environmental effects

1. Introduction

To better map the folding/unfolding mechanisms of β -structure in proteins, various β -hairpin model systems have been studied with both experimental and computational methods (Lewandowska et al., 2010; Schaeffer et al., 2008), including the double hairpin WW domain, which comprises a stable, soluble 3-stranded β -sheet (Fig. 1A). Experimental Φ -value analysis of PIN1 (Jager et al., 2001) and FBP28 (Petrovich et al., 2006) WW domains revealed high Φ -values in the turn of the first hairpin. High Φ -values suggest that the residues being probed are structured in the transition state (TS). Conversely low values suggest a lack of structure at the site probed. A similar pattern of Φ -values was seen in studies of backbone amide bonds substituted with thio-ether bonds in hYAP (Ferguson et al., 2001) or ester bonds in PIN1 (Deechongkit et al., 2004). The PIN1 WW domain results were used to argue that the first β -hairpin folds via a zipper mechanism, with folding nucleated by

*Corresponding author: daggett@u.washington.edu.

the first turn (Jager et al., 2001; Nguyen et al., 2003). However, NMR and molecular dynamics (MD) simulations of PIN1, hYAP and FBP28 WW domains indicated that the first turn is flexible even in the native state (Sharpe et al., 2007). All-atom MD unfolding simulations showed good agreement between the MD-generated TS structures and the experimental Φ -values, allowing for the examination of the unfolding simulation in detail to study the unfolding/folding pathway (Petrovich et al., 2006). The backbone hydrogen bonds between the first two strands of FBP28 WW domain did not form in any discernable pattern in the simulations, contrary to what would be expected for a zipper mechanism for β -hairpin folding. Instead, the first turn was present as a kink in the backbone that allowed the side chains of the residues in the strands to form contacts, leading to hydrogen bond formation between the strands (Petrovich et al., 2006; Sharpe et al., 2007).

But one wonders whether the FBP28 WW domain, or any β -hairpin peptide, is a realistic model of the folding of β -structure in the context of a larger protein with a conventional hydrophobic core. To investigate this question we began with the double hairpin motif of the WW domain and searched for proteins containing this motif in our Dynameomics database of simulations (van der Kamp et al., 2010). The Dynameomics targets are chosen from our Consensus Domain Dictionary (Schaeffer et al., 2011), in which all autonomous protein domains are divided into a set of protein metafolds. The resulting 807 metafolds are then ranked based on population, such that rank 1 has the highest population (Schaeffer et al., 2011). The Dynameomics database contains simulations, in explicit water, from at least one member of each metafold (van der Kamp et al., 2010), including the FBP28 WW domain (rank 194). Three good protein candidates containing the WW domain's double hairpin motif were identified: cold shock protein A (CspA) and cold shock protein B (CspB) from metafold rank 9 and the IIA domain from glucose permease from metafold rank 352 (Fig. 1A).

The two cold shock proteins, CspA from *E. coli* and CspB from *Bacillus subtilis*, have a double hairpin comprised of their first three β -strands (Fig. 1A,B). Folding studies of CspA have focused on the role of the solvent exposed hydrophobic patch, which is important for nucleic acid binding (Garcia-Mira and Schmid, 2006; Reid et al., 1998; Rodriguez et al., 2000). These studies suggest that the TS is native-like and CspA folds via a two-state mechanism. One method to quantify the amount of structure present at the residue level in the protein folding/unfolding TS is Φ -value analysis (Fersht et al., 1987; Matouschek et al., 1989). This method relies on introducing mutations and evaluating the effect on stability and the rate of folding/unfolding relative to the wild-type protein. A ratio of the effect on the rate versus stability generally take values between 0 and 1, from unstructured to native-like extent of structure at the mutation site, respectively. Φ -value analysis has been performed on CspB (Garcia-Mira et al., 2004; Garcia-Mira and Schmid, 2006), and the highest Φ -values are in the first strand (Garcia-Mira et al., 2004).

The third and largest protein in this study, glucose permease IIA domain, also has a double hairpin similar to those of the WW domain and the cold shock proteins (1.8 Å C α -RMSD between strand residues), but the hairpins are part of a larger β -sheet (Fig. 1). The β 4- β 5- β 6 double hairpin lacks the aromatic residues present in both the cold shock proteins and FBP28 WW domain (Fig. 1B). Here, we examine the major unfolding pathways found in both of the cold shock proteins and in the IIA domain of glucose permease. We compare and contrast these protein simulations to the unfolding pathways of the FBP28 WW domain and the isolated motifs excised from each protein, noting that we present representative results from 1.2 μ s of simulation time sampled for this study.

2. Methods

One 298 K simulation and at least five 498 K unfolding simulations were performed of CspA (PDB code: 1mjc) (Schindelin et al., 1994), CspB (PDB code: 1csp) (Schindelin et al., 1993) and glucose permease IIA domain (PDB code: 1gpr) (Herzberg, 1992). In addition to the previously published FBP28 WW domain simulations (Petrovich et al., 2006; Sharpe et al., 2007), a new 298 K and five 498 K unfolding simulations (PDB code: 1e0l) (Macias et al., 2000) were performed using the Dynameomics protocol for direct comparison with the protein simulations (Beck et al., 2008; Van der Kamp et al., 2010). Simulations were performed using the *in lucem* molecular mechanics package (*ilmm*) (Beck et al., 2000–2011) and the protein and water potential functions of Levitt *et al.* (Levitt et al., 1995, 1997). The microcanonical ensemble (constant number of particles, volume, and energy) and periodic boundary conditions were employed. Each protein was subjected to 1000 steps of steepest decent minimization and then solvated in a box of water with the appropriate experimental density of 0.997 g/ml for 298 K (Kell, 1967) and 0.829 g/ml for 498 K (Haar et al., 1984). The water and protein were equilibrated by 500 steps of minimization of the water, followed by 1 ps of dynamics of the water molecules alone. Finally, 500 steps of energy minimization of the water and 500 steps of minimization of the protein atoms were performed. All minimization and MD calculations were unrestrained. A 2 fs time step was used in each simulation and other details regarding the simulations have been presented (Beck and Daggett, 2004). The native state simulations used a 10 Å force-shifted cutoff for nonbonded interactions (Beck et al., 2005; Levitt et al., 1995), while the unfolding simulations used an 8 Å cutoff. Past studies suggest that an 8 Å is adequate at high temperature because of the expansion of the structures (Beck et al., 2005), but the longer cutoff is preferred for the more tightly packed native state at lower temperature, as it is more sensitive to the cutoff. The 298 K simulations and at least two of the 498 K simulations were at least 51 ns. The additional 3–5 unfolding simulations were short (at least 2 ns each) to further sample the early events of unfolding in more detail. Structures were saved every 1 ps for analysis for the longer simulations except for the first 2 ns, where they were saved every 0.2 ps. For the shorter (2–5 ns) 498 K simulations the structures were saved every 0.2 ps. Overall, 66 simulations were performed totaling 1206 ns, or 1.2 μs, of simulation time. See supplementary Table S1 for a breakdown of the simulations performed and simulation times.

Simulations of the excised double hairpin motifs from CspA, CspB and IIA domain were also performed. The double hairpin motif from each protein was excised from the original PDB structure of each protein: residues 2–36 for CspA, 1–32 for CspB and 55–86 for IIA domain. The sequences for each double hairpin motif used in the excised simulations are shown in Fig. 1B. We performed 6 simulations of each excised motif using the same simulation protocol was used as described above. We ran a control simulation at 298 K for 21 ns and 5 unfolding simulations at 498 K for 5 ns on each excised motif.

We identified TS ensembles from the unfolding simulations using our standard conformational clustering method (Li and Daggett, 1994, 1996). The C α -RMSD was calculated between each pair of structures in a given simulation. We then used multidimensional scaling to reduce the matrix a 3D representation to identify clusters of similar structures. The 5 ps window of structures before the exit from the native-like cluster was assigned as the TS ensemble for that simulation. This corresponds to 26 structures in each TS ensemble from each unfolding simulation.

In order to quantify the amount of structure present in the TS, we calculated S values, or structure indices, for each residue in the protein across the TS ensemble (Daggett et al., 1996). S values are a product of two terms S_3° and S_2° . S_3° is the ratio of the number of

contacts present in the TS ensemble to the number of contacts in the native state. S_2° is the fraction of native secondary structure present in the TS.

3. Results and Discussion

3.1 WW Domain

Our published simulations of FBP28 WW domain were performed at low pH at 333, 348 and 373 K (Petrovich et al., 2006; Sharpe et al., 2007). Additional unfolding simulations of the FBP28 WW domain have been performed at 498 K at neutral pH as part of our Dynameomics project (Beck et al., 2008; Van der Kamp et al., 2010). The unfolding pathway of FBP28 WW domain did not change over the 333 – 373 K temperature range (Petrovich et al., 2006). The new 498 K neutral pH results presented here (Fig. 2) indicate that the pathway is the same even at very elevated temperatures, as observed for other proteins (Day et al., 2002; DeMarco et al., 2004; Ferguson et al., 2005; Mayor et al., 2000). The third strand was the first to unfold in the major unfolding pathway at all temperatures (Fig. 2A, B), which is consistent with experiments probing unfolding and refolding (Petrovich et al., 2006). Furthermore, unfolding has been shown to be the reverse of folding as demonstrated by comparison with experiment and by simulations of proteins at their melting temperatures with continuous trajectories of unfolding and refolding (Day and Daggett, 2007; McCully et al., 2008).

In all cases, the backbone was very flexible in the first turn of FBP28. Since this turn is flexible even in the native state (Sharpe et al., 2007), it's not surprising that the backbone angles fluctuate during unfolding. These fluctuations artificially lowered the S-values, which reflect the combined extent of secondary and tertiary structure on a per-residue basis for comparison with experimentally derived Φ -values (Daggett et al., 1996), in the turn even in the presence of side chain interactions. Therefore, the tertiary component of the S-values better represents the amount of structure in the turn in the TS ensemble (Petrovich et al., 2006). The resulting S and Φ -values, which probe site-specific structure experimentally, are in good agreement for the FBP28 WW domain, independent of temperature and pH ($R = 0.77$ between S and Φ -values) (Petrovich et al., 2006).

As the WW domain unfolded, there was no consistent order to the loss of hydrogen bonds in either hairpin (Petrovich et al., 2006; Sharpe et al., 2007). The same results were found in the new 498 K simulations presented here (Fig. 2C), even as the hydrogen bonds were lost very quickly due to the rapid unfolding induced by the very high temperature. After the backbone hydrogen bonds were lost, there were long-lasting side chain contacts in the simulations between large hydrophobic residues. Examining the unfolding simulations of the WW domain in reverse to reflect the folding pathway revealed that the residues in the first turn formed a kink, allowing side chain contacts to form between residues in the first two strands. These contacts brought the backbones into contact and allowed for the formation of hydrogen bonds, in no consistent order, to form the first hairpin. The first turn acted as a diffuse nucleus to facilitate the coalescence of residues in the first two strands. But, unlike the zipper model of hairpin formation, the precise structure of the first turn did not drive folding (Petrovich et al., 2006; Sharpe et al., 2007).

3.2 Cold Shock Proteins

The first event in each of the unfolding simulations of CspA was the first three strands comprising the WW double hairpin motif ($\beta 1$ – $\beta 3$) separated from the core of the protein (Fig. 3A, 3B). As these strands moved away from the core, contacts were lost between $\beta 2$ and $\beta 3$ (Fig. 3A,C), similar to the early loss of structure in $\beta 3$ in the FBP28 WW domain. The TS structure identified from the unfolding simulations showed more contacts between the

residues in the first two strands than in $\beta 3$. The TS ensemble was heterogeneous with an average C α RMSD between TS structures within and across different simulations of 4.5 ± 0.5 Å.

Similar to the WW domain, CspA has aromatic residues in the first three strands including three Phe residues that are important to nucleic acid binding and protein function. Experimental Φ -value analysis was performed on these 3 residues (Rodriguez et al., 2000). Two mutations were made to Phe18, one to Ser and another to Leu, and the resulting Φ -values do not agree (0.04 for F18L and 0.64 for F18S), making it difficult to estimate the amount of structure present in the TS at this position from experiment. The average MD-derived S value of Phe18 falls between the two experimentally derived Φ -values reported at this position. The other two positions tested, Phe20 and Phe31 gave large negative Φ -values for mutation to Leu, falling outside of the normal 0–1 range and suggesting that the mutations are not perturbations probing wild-type behavior, which is a fundamental assumption of Φ -value analysis. The most reliable Φ -values tend to involve conservative hydrophobic core mutations, particularly those deleting methylene groups (Otzen et al., 1994).

Unlike WW domains, CspA doesn't have a backbone hydrogen bond in the first turn but monitoring the loss of hydrogen bonds between the first two strands showed no discernable pattern (Fig. 3C). In 4 of the 8 unfolding simulations of CspA there was a loss of the hydrogen bonds in the middle of the first hairpin before the outer hydrogen bonds and in two simulations all hydrogen bonds were lost between the strands at approximately the same time. This finding supports the conclusion from our WW domain studies (Petrovich et al., 2006; Sharpe et al., 2007) that the β -hairpins do not fold via an ordered, zipper mechanism.

The unfolding pathway of CspB was very similar to CspA. The first event in unfolding was expansion of the protein, pulling the first three strands away from the rest of the protein. Next, structure was lost in the C-terminal strands ($\beta 4$ and $\beta 5$, grey in Fig. 3D), followed by loss of structure in the first three strands, the double hairpin motif. In 4 of the 5 CspB unfolding simulations the first strand shortened, starting at the N-terminus as contacts were lost between $\beta 1$ and $\beta 2$ (red and green in Fig. 3D). However, contacts were maintained between the C-terminal half of strand 1 and strand 2. The third strand then fluctuated away from the first hairpin. Contacts between residues in the first two strands were present late into the unfolding process. After the first ns the native structure was completely lost.

TS ensembles were identified for the CspB unfolding simulations and S values were calculated. The TS ensemble across and within simulations is more heterogeneous than that of CspA with an average C α RMSD between structures of 6.1 ± 0.8 Å. The results of a Φ -value analysis are available for CspB, for which there are 16 mutants with a sufficient change in free energy of folding upon mutation ($\Delta\Delta G_{D-N}$ above 0.7 kcal/mol, 2.9 kJ/mol) (Garcia-Mira et al., 2004). The correlation coefficient between the average S values over all 5 unfolding simulations and the experimental Φ -values was 0.61. The TS ensemble identified from the second unfolding simulation showed very poor agreement with experiment ($R=0.19$). This simulation lost structure slowly, taking over 5 ns to lose the native hydrogen bonds between strands $\beta 4$ and $\beta 5$, making TS assignment difficult. The other simulations unfolded much more rapidly, typically losing hydrogen bonds between the strands in the first couple of ns. When simulation #2 was removed and the average S values recalculated, the correlation with experiment increased ($R=0.70$).

Two residues are in particularly poor agreement. Residues Val26 and Leu41 both have lower average S values (0.24 ± 0.15 and 0.03 ± 0.03 , respectively) than experimental Φ -values (0.74 for V26T and 0.40 for L41A). Val26 is part of strand $\beta 3$ and is buried in the

starting structure. In all unfolding simulations, Val26 becomes solvent exposed, losing tertiary contacts, early in unfolding. The V26T mutation is problematic due to the change in polarity; more conservative mutations are generally desired with Φ -value analysis. Leu41 is located in a loop region and is part of the protein core, which makes contacts to the double hairpin motif. Leu41 also becomes solvent exposed early in the unfolding pathway, as the double hairpin motif loses contact with the protein core. As a result, the average S -value is low. When these two problematic residues are removed from consideration, the correlation between the remaining 14 pairs of average S and Φ -values was 0.80 (again without simulation #2, Fig. 4). Assuming that these Φ -values are correct and there is no change in pathway upon mutation, the simulations are either underestimating the extent of structure for these residues or neglecting solvation effects is giving rise to the discrepancies.

Similar to the unfolding pathway seen in CspA and the FBP28 WW domain, there was no consistent order to the loss of hydrogen bonds between the strands in the β -hairpins of CspB. Also in agreement with the CspA unfolding simulations, all hydrogen bonds between $\beta 2$ and $\beta 3$ were lost quickly and at approximately the same time in the unfolding pathway. For both hairpins the main chain and side chain contacts were lost together, without the persistent side chain contacts found in the FBP28 WW domain simulations.

3.3 Glucose Permease IIA Domain

Unlike CspA, CspB and FBP28 WW domain, there was no consistent loss of the third strand in the double hairpin motif ($\beta 6$) before the first hairpin ($\beta 4$ – $\beta 5$) (Fig. 5A). Instead, in 4 of the 8 unfolding simulations the hydrogen bonds in the second hairpin ($\beta 5$ – $\beta 6$) broke either at the same time or before the first hairpin. In all 8 simulations, the two hydrogen bonds between residues Asn62 and Gly71 were the first interactions lost in the double hairpin motif (Fig. 5B). The C-terminal end of $\beta 4$, near the hairpin turn, bulged up and away from $\beta 5$. In half of the simulations contacts were regained between the C-terminal end of $\beta 4$ and the start of $\beta 5$ as contacts were lost to $\beta 6$. Therefore, similar to the cold shock proteins and to FBP28 WW domain, there was no evidence of a zipper-like unfolding mechanism in the first hairpin and the precise structure of the turn was lost early in the unfolding pathway.

As mentioned previously, we did not see a consistent loss of $\beta 5$ – $\beta 6$ contacts before $\beta 4$ – $\beta 5$ unfolds. There was, however, a consistent loss of the attached strands (non-WW motif) in the β -sheet early in unfolding (strands $\beta 10$, $\beta 2$ and $\beta 1$, see Figs. 1 and 5A). Unlike the cold shock protein simulations, the double hairpin motif did not separate from the rest of the protein as it unfolded (Fig. 5C). Instead, there was initially a decrease in native contacts between the motif and the rest of the protein as the core expanded; however, the motif maintained ~50% of the extent of the starting contacts to the rest of the protein throughout the early unfolding events, which amounted to approximately 3 times the contacts in the cold shock proteins (Fig. 5C). Unfortunately experimental results for this protein are not available for comparison.

There is a salt bridge network before the start of $\beta 4$ involving Glu4 (the first residue), Arg57 (loop residue immediately before $\beta 4$), Asp75 (in $\beta 5$ – $\beta 6$ turn) and Arg105 ($\beta 8$). At the beginning of the simulations, Lys59 (near the N-terminus of $\beta 4$) is oriented away from this salt bridge network. Instead, Lys59 interacted with the backbone of Glu102 (residue immediately following $\beta 7$). In the majority of the unfolding pathways, Lys59 swung around to join the salt bridge network. As other interactions kept the $\beta 4$ – $\beta 5$ turn close to the $\beta 7$ region of the protein (such as a nonnative salt bridge between Lys67 of the $\beta 4$ – $\beta 5$ turn and Glu94 immediately before $\beta 7$), the change in orientation of the Lys59 side chain caused strain in the $\beta 4$ – $\beta 5$ hairpin (Fig. 6).

3.4 Excised Double Hairpin Motifs

In addition to comparing the unfolding of the double hairpin motif in the context of the full proteins and the FBP28 WW domain, we also performed simulations of the excised motifs in water. The unfolding pathways of the excised CspA and CspB double hairpin motifs were very similar to the unfolding pathways within the full-length proteins, which in turn is what was observed for FBP28. In all cases, contacts to the third strand in the motif ($\beta 3$) were lost before the first hairpin ($\beta 1-2$) unfolded, as observed for FBP28 (Fig. 7A,B,C). This was expected as the double hairpin motif pulled away from the core of the protein early in the CspA and CspB simulations (Fig. 4A,D), in effect isolating the motif.

In contrast, the double hairpin motif in the IIA domain (strands $\beta 4-5-6$), maintained contacts to the rest of the protein during the unfolding simulations of the full-length protein. As a result, the unfolding of the motif itself was more varied when compared to the cold shock proteins or the FBP28 WW domain unfolding pathways. The unfolding pathway of the double hairpin motif of the IIA domain was influenced by the salt bridge network near the start of $\beta 4$, as discussed above. In the starting structure of the excised IIA domain double hairpin motif, this salt bridge network is smaller and the salt bridge between Arg57 and Asp75 is maintained. As seen in the full-length IIA simulations, the side chain of Lys59 swings around to join the salt bridge network. However, in the excised domain simulations, the first hairpin of the motif is flexible enough to accommodate this movement, resulting in the same unfolding pathway as observed in the other excised motifs, FBP28, and the cold shock proteins (Fig. 7D).

4. Conclusions

The unfolding simulations of the two different cold shock proteins (CspA and CspB) showed a consistent unfolding pathway with early loss of contacts between the first three β -strands (the double hairpin motif, or WW domain motif) and the hydrophobic core thereby isolating the double hairpin motif. The proteins then lost structure in $\beta 3$ similar to the FBP28 WW domain unfolding simulations. The same unfolding pathway was seen in simulations of the excised double hairpin motif from the cold shock proteins (Fig. 7A,B and C). Thus, for these proteins the folding behavior of the model system is representative of the behavior of that same motif in the protein independent of sequence. While the sequence identity between FBP28 and the double hairpin motifs in CspA and CspB is low (Fig. 1B), they are similar in the sense that they have at least two aromatic residues and polar residues in both $\beta 1$ and $\beta 2$. In contrast, $\beta 3$ is very polar in FBP28 while it is quite hydrophobic in CspA and CspB.

This pattern was not repeated in the unfolding simulations of the larger, more complex protein glucose permease IIA domain (Fig. 5). For the latter system contacts to the hydrophobic core were both more numerous and persistent over time, which led to a different order of structural loss depending on strength of interactions between the motif and the core. However, without the contacts anchoring the double hairpin motif to the rest of the protein, the first hairpin in the double hairpin motif behaved as in the WW domain. Consequently, in the case of the IIA domain, context affects the folding of the motif such that the intrinsic unfolding pathway involves early loss of $\beta 3$ is not followed due to tertiary interactions with the rest of the protein. The intrinsic unfolding behavior was preserved for all sequences when the motif was excised from the rest of the protein.

In all proteins studied, there was no order to the loss of hydrogen bonds between the β -strands, indicating that these hairpins do not unfold via an ordered zipper mechanism. Instead the precise structure of the turns was lost early in the unfolding process in agreement with the results from the FBP28 WW domain studies. In fact, in the case of the IIA domain the hydrogen bond next to the turn of the first hairpin was the first hydrogen bond in the

sheet to break. In addition, a significant difference between the FBP28 WW domain unfolding pathway and those of the larger β -proteins was the lack of persistent side chain contacts after the loss of main chain contacts. Therefore, while the hairpins shared similar unfolding behavior in terms of secondary structure loss, the presence of the surrounding protein altered the details of the pathway in each case. In the case of the Csp proteins the effects were relatively minor and the pathways were generally similar, but in the case of the glucose permease IIA domain the unfolding pathways of the double hairpin motif were different (compare Figs. 2, 3, and 5). Consequently, the context in which a structural motif resides is critical in determining its behavior, i.e. the contacts to the other regions of the protein affect the unfolding pathway of the double hairpin motif and of course sequence differences affect the contacts. This is consistent with our earlier findings that tertiary contacts influence the formation of secondary structure, both helix and β -structure (Bond et al., 1997; Wong et al., 2000; Scott et al., 2006; Alonso et al., 2001); here the effect is observed on unfolding instead of refolding.

WW domains have been used as models for the general folding behavior of β -sheets. Here we have taken it a step further: we compare the behavior of the isolated motif with that same motif in different proteins. Interactions between the motif and the surrounding protein residues versus just solvent allows for a description of the intrinsic folding/unfolding of structural motifs and perturbations to this intrinsic, or model behavior, due to tertiary contacts. In the case of the cold shock proteins the perturbation is minimal with respect to the unfolding pathway, but in the case of the glucose permease IIA domain the behavior of the isolated motif is a poor model for that same structure in the context of the protein independent of whether the isolated motif is the natural FBP28 WW domain or the sequence-matched excised motif from the IIA domain. Our results emphasize the need for caution when interpreting or extrapolating the results from model systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for financial support provided by the National Institutes of Health (GM 50789). We are also grateful for support from the External Research Program of Microsoft Research for development of our Dynameomics database. The Dynameomics trajectories were performed using resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Alonso DOV, DeArmond S, Cohen FE, Daggett V. Mapping the early steps of the pH-induced conformational conversion of the prion protein. *Proc Natl Acad Sci USA*. 2001; 98:2985–2989. [PubMed: 11248018]
- Beck, DAC.; Alonso, DOV.; Daggett, V. *in lucem* Molecular Mechanics (*ilmm*). University of Washington; Seattle: 2000–2011.
- Beck DAC, Daggett V. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods Enzymol*. 2004; 34:112–120.
- Beck DAC, Armen RS, Daggett V. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry*. 2005; 44:609–616. [PubMed: 15641786]
- Beck DAC, Jonsson AL, Schaeffer RD, Scott KA, Day R, et al. Dynameomics: Mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Eng, Des Sel*. 2008; 21:353–368. [PubMed: 18411224]

- Bond CJ, Wong K, Clarke J, Fersht AR, Daggett V. Characterization of Residual Structure in the Thermally Denatured State of Barnase by Simulation and Experiment: Description of the Folding Pathway. *Proc Natl Acad Sci USA*. 1997; 94:13409–13413. [PubMed: 9391038]
- Daggett V, Li AJ, Itzhaki LS, Otzen DE, Fersht AR. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol*. 1996; 257:430–440. [PubMed: 8609634]
- Day R, Bennion BJ, Ham S, Daggett V. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J Mol Biol*. 2002; 322:189–203. [PubMed: 12215424]
- Day R, Daggett V. Direct Observation of Microscopic Reversibility in Single-molecule Protein Folding. *J Mol Biol*. 2008; 366:677–686. [PubMed: 17174331]
- Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, et al. Context-dependent contributions of backbone hydrogen bonding to β -sheet folding energetics. *Nature*. 2004; 430:101–105. [PubMed: 15229605]
- DeMarco ML, Alonso DOV, Daggett V. Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein. *J Mol Biol*. 2004; 341:1109–1124. [PubMed: 15328620]
- Ferguson N, Pires JR, Toepert F, Johnson CM, Pan YP, et al. Using flexible loop mimetics to extend Φ -value analysis to secondary structure interactions. *Proc Natl Acad Sci USA*. 2001; 98:13008–13013. [PubMed: 11687614]
- Ferguson N, Day R, Johnson CM, Allen MD, Daggett V, Fersht AR. Simulation and experiment at high temperatures: Ultrafast folding of a thermophilic protein by nucleation-condensation. *J Mol Biol*. 2005; 347:855–870. [PubMed: 15769475]
- Fersht AR, Leatherbarrow RJ, Wells TN. Structure-activity relationships in engineered proteins: Analysis of use of binding energy by linear free energy relationships. *Biochemistry*. 1987; 26:6030–6038. [PubMed: 3480005]
- Garcia-Mira MM, Boehringer D, Schmid FX. The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol*. 2004; 339:555–569. [PubMed: 15147842]
- Garcia-Mira MM, Schmid FX. Key role of Coulombic interactions for the folding transition state of the cold shock protein. *J Mol Biol*. 2006; 364:458–468. [PubMed: 17020767]
- Haar, L.; Gallagher, JS.; Kell, GS. NBS/NRC steam tables: thermodynamic and transport properties and computer programs for vapor and liquid states of water in SI units. Hemisphere; Washington, DC: 1984.
- Herzberg O. An atomic model for protein-protein phosphoryl group transfer. *J Biol Chem*. 1992; 267:24819–24823. [PubMed: 1447219]
- Jager 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]
- Kell GS. Precise representation of volume properties of water at one atmosphere. *J Chem Eng Data*. 1967; 12:66–68.
- Levitt M, Hirshberg M, Sharon R, Daggett V. Potential-Energy function and parameters for simulations of the molecular-dynamics of proteins and nucleic-acids in solution. *Comput Phys Commun*. 1995; 91:215–231.
- Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J Phys Chem B*. 1997; 101:5051–5061.
- Lewandowska A, Oldziej S, Liwo A, Scheraga HA. β -hairpin-forming peptides; models of early stages of protein folding. *Biophys Chem*. 2010; 151:1–9. [PubMed: 20494507]
- Li AJ, Daggett V. Characterization of the transition-state of protein unfolding by use of molecular-dynamics - chymotrypsin inhibitor-2. *Proc Natl Acad Sci USA*. 1994; 91:10430–10434. [PubMed: 7937969]
- Li AJ, Daggett V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J Mol Biol*. 1996; 257:412–429. [PubMed: 8609633]
- Macias MJ, Gervais V, Civera C, Oschkinat H. Structural analysis of WW domains and design of a WW prototype. *Nat Struct Biol*. 2000; 7:375–9. [PubMed: 10802733]

- Matouschek A, Kellis JT, Serrano L, Fersht AR. Mapping the transition-state and pathway of protein folding by protein engineering. *Nature*. 1989; 340:122–126. [PubMed: 2739734]
- Mayor U, Johnson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci USA*. 2000; 97:13518–13522. [PubMed: 11087839]
- McCully ME, Beck DAC, Daggett V. Microscopic Reversibility of Protein Folding in Molecular Dynamics Simulations of the Engrailed Homeodomain. *Biochemistry*. 2008; 47:7079–7089. [PubMed: 18553935]
- 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]
- Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci USA*. 1994; 91:10422–10425. [PubMed: 7937967]
- 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]
- Reid KL, Rodriguez HM, Hillier BJ, Gregoret LM. Stability and folding properties of a model β -sheet protein, *Escherichia coli* CspA. *Protein Sci*. 1998; 7:470–479. [PubMed: 9521124]
- Rodriguez HM, Vu DM, Gregoret LM. Role of a solvent-exposed aromatic cluster in the folding of *Escherichia coli* CspA. *Protein Sci*. 2000; 9:1993–2000. [PubMed: 11106173]
- Schaeffer RD, Fersht AR, Daggett V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr Opin Struct Biol*. 2008; 18:4–9. [PubMed: 18242977]
- Schaeffer RD, Jonsson AL, Simms AM, Daggett V. Generation of a consensus protein domain dictionary. *Bioinformatics*. 2011; 27:46–54. [PubMed: 21068000]
- Schindelin H, Marahiel MA, Heinemann U. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature*. 1993; 364:164–168. [PubMed: 8321288]
- Schindelin H, Jiang W, Inouye M, Heinemann U. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc Natl Acad Sci USA*. 1994; 91:5119–5123. [PubMed: 8197194]
- Scott KA, Alonso DOV, Pan Y, Daggett V. Importance of Context in Protein Folding: Secondary Structural Propensities versus Tertiary Contact-Assisted Secondary Structure Formation. *Biochemistry*. 2006; 45:4153–4163. [PubMed: 16566589]
- Sharpe T, Jonsson AL, Rutherford TJ, Daggett V, Fersht AR. The role of the turn in β -hairpin formation during WW domain folding. *Protein Sci*. 2007; 16:2233–2239. [PubMed: 17766370]
- Van der Kamp MW, Schaeffer RD, Jonsson AL, Scouras AD, Simms AM, et al. Dynameomics: A comprehensive database of protein dynamics. *Structure*. 2010; 18:423–35. [PubMed: 20399180]
- Wong K, Clarke J, Bond CJ, Neira JL, Freund SMV, et al. Towards Complete Characterization of the Structural and Dynamic Properties of the Denatured State of Barnase and the Role of Residual Structure in Folding. *J Mol Biol*. 2000; 296:1257–1285. [PubMed: 10698632]

Fig. 1. Double hairpin motif

A) Structures of FBP28 WW domain, CspA, CspB and glucose permease IIA domain with the double hairpin motif colored in red, green and blue. The extra strands in the β -sheet of glucose permease IIA domain are also labeled. For FBP28 WW domain residues 8–12 are red, 18–22 green, and 26–29 are blue for β 1, β 2 and β 3, respectively. CspA is colored red for β 1 residues 5–13, green for β 2 residues 18–23, and blue for β 3 residues 30–33. For CspB residues 2–10 are red for β 1, 15–19 green for β 2 and 26–29 blue for β 3. Glucose permease IIA domain is colored red for β 4 residues 58–63, green for β 5 residues 69–74, and blue for β 6 residues 79–83. The same coloring is used throughout the figures. B) Sequence alignment of the double hairpin motif in each system, with arrows marking the β -strands.

(A) FBP28 Unfolding Pathway at 373 K

(B) FBP28 Unfolding Pathway at 498 K

(C) β 1-2 Hydrogen Bond Loss at 498 K **Fig. 2. WW Domain Unfolding Pathway**

A) Snapshots from a 373 K, low pH unfolding simulation. B) Snapshots from a 498 K, neutral pH unfolding simulation. C) The presence of hydrogen bonds in the first β -hairpin over time in the same simulation as shown in B. If a hydrogen bond is present between the specified atoms, a dot is graphed.

(A) CspA Unfolding Pathway

(B) Double Hairpin Motif Contacts Over Time

(C) β 1-2 Hydrogen Bond Loss

(D) CspB Unfolding Pathway

Fig. 3. Cold Shock Protein Unfolding Pathway

A) Snapshots from simulation #1 of the CspA 498 K unfolding simulation. B) Hydrogen bonds between strands β 1 and β 2 over time for the simulation shown in panel A. C) Total number of atom-atom contacts over time for the simulation in panel A. In gray are contacts between the residues in the double hairpin motif and the rest of the protein. In black are contacts between just the residues in the double hairpin motif (strands β 1–3). D) Snapshots from simulation #1 498 K CspB unfolding simulation.

Fig. 4. Comparison of the experimental Φ -values and calculated S-values for the TS CspB
 A representative structure from the TS ensemble is shown in Fig. 3D. The S-values were calculated over all structures in the TS ensemble pooled from all simulations. Points cover the 14 mutant set described in the text.

(A) Glucose Permease IIA Domain Unfolding Pathway

(B) β 4-5 Hydrogen Bond Loss

(C) Double Hairpin Motif Contacts Over Time

Fig. 5. Glucose Permease IIA Domain Unfolding Pathway

A) Snapshots from simulation #1 of the IIA domain 498 K unfolding simulation. Note that the TS structure is representative of the TS ensemble, which has an average $C\alpha$ RMSD within and between different simulations of 5.8 ± 0.8 Å. B) Hydrogen bonds between strands β 4 and β 5 over time for the simulation shown in panel A. C) Total number of atom-atom contacts over time for the simulation in panel A. In gray are contacts between the residues in the double hairpin motif and the rest of the protein. In black are contacts between just the residues in the double hairpin motif (strands β 4–6).

(A) Glucose permease IIA domain unfolding pathway

(B) Unfolding pathway of double hairpin motif excised from IIA domain

Fig. 6. Comparison of the Thermal Unfolding of the Double Hairpin Motif in Glucose Permease IIA Domain with the Excised Motif

A) The unfolding of the double hairpin motif in the context of the full IIA domain with the main side chains of residues involved in perturbing the motif's unfolding pathway displayed and labeled. B) The unfolding of the excised double hairpin motif from the IIA domain.

Note the similarities between the unfolding of this domain and the WW domain in Figure 2.

Fig. 7. Comparison of the Thermal Unfolding of the isolate Double Hairpin Motif

A) FBP28 WW domain. B) Motif excised from CspA. C) Motif excised from CspB. D) Motif excised from glucose permease IIA domain. All simulations are at 498 K. Note the similarities in the unfolding of the secondary structure across all isolated domains despite differences in their sequences.