Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Caitlin Mearns Marlatt Davis Date

Heterogeneity in fast-folding beta proteins

By

Caitlin Mearns Marlatt Davis
Doctor of Philosophy

Chemistry

_____

R. Brian Dyer, Ph.D.
Advisor

_____

David G. Lynn, Ph.D.
Committee Member

_____

James T. Kindt, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Heterogeneity in fast-folding beta proteins

By

Caitlin Mearns Marlatt Davis
B.S., University of Michigan, 2007

Advisor: R. Brian Dyer, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2014

Abstract

Heterogeneity in fast-folding beta proteins
By Caitlin Mearns Marlatt Davis

Most proteins self-assemble, so secondary structure and folding dynamics must be encoded in the protein's sequence. Amyloids, extended β-sheet structures, have been associated with protein misfolding diseases such as Alzheimer's and prion disease. Understanding the secondary structure of these β-sheet structural motifs and the rules that govern their folding will ultimately allow for the design of more effective treatments of these diseases. Of particular interest are the early kinetic events, which structures form first and on what timescale. Laser-induced temperature-jumps coupled with fluorescence or infrared spectroscopy have been used to probe changes in the peptide backbone on the submillisecond timescale. Wavelength dependent infrared measurements allow resolution of the folding mechanism by independently probing the dynamics associated with the β-sheet or β-turn. In this dissertation, a hierarchal approach was taken to study β-sheet structures, starting from the smallest β-sheet structural motif, the β-hairpin, and building up to more complex systems, extended β-sheets and membrane bound β-hairpins. Even for the fastest folding linear β-hairpin, CLN025, folding is heterogeneous and cannot be described by a simple two state model. This work was extended to systems with two β-hairpins, WW domains, which increased the folding complexity. In all WW domains studied, folding was initiated in the turns and the β-sheets form last. Replacing the slower folding β-hairpin with a faster one increased the WW Domain folding speed by an order of magnitude, demonstrating the importance of β-hairpins in the context of the larger domain. β-hairpin formation was also shown to be an early step in the mechanism of membrane-induced folding and insertion of cationic anticancer β-hairpin therapeutics. These studies demonstrate that small fast folding domains are relevant to understanding larger systems, and further, while the folding of these systems is heterogeneous and cannot be described by a simple two-state model there are general rules that govern the folding of all of these systems.

Heterogeneity in fast-folding beta proteins

By

Caitlin Mearns Marlatt Davis
B.S., University of Michigan, 2007

Advisor: R. Brian Dyer, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2014

Acknowledgement

I would like to thank Dr. Brian Dyer for his assistance, advice and support throughout the course of my Ph.D. research.

I would also like to thank Dr. James Kindt for his guidance and patience throughout our collaboration.

Thank you to Dr. Dan Raleigh for supplying the CLN025 peptide.

Many thanks to the members of the Dyer lab, past and present, for their company and support.

I would like to thank my lab colleagues on the folding project Dr. Suresh Nagarajan, Kat Cooper, Erin Schuler, Sam Jeong and Gokul Raghunath.

A large debt of gratitude is owed to Michael Reddish for his assistance with all manner of lab difficulties and his friendship over the years.

Special thanks to my mentee, Daniela Ruiz, for challenging me to grow as a mentor.

Thank you to Dr. Leah Roesch, Dr. Pat Marsteller and Dr. Jacob Shreckengost for providing me with opportunities to mentor students through the Scholarly Inquiry and Research at Emory (SIRE) and Clare Boothe Luce (CBL) Undergraduate Research Programs.

Thank you to my husband, Dr. Mike Davis, for showing me that getting a Ph.D. really is that easy.

Finally, thank you to my parents, Judd and Fran Marlatt, for their love and for inspiring me to always work hard and set ambitious goals.

Table of Contents

List of Figures

Chapter 1 Figures

Chapter 2 Figures

Chapter 3 Figures

Chapter 4 Figures

Chapter 5 Figures

Chapter 6 Figures

Chapter 7 Figures

Appendix I Figures

List of Tables

Chapter 5 Tables

Chapter 6 Tables

Chapter 7 Tables

Appendix I Tables

Chapter 1: Introduction

1.1 The Protein Folding Problem

Proteins are essential functional elements in the cell.[1] They have a wide range of functions including catalyzing reactions, replicating DNA, responding to stimuli and transporting molecules. For a protein to function properly, it must be folded into the correct three-dimensional shape. Most proteins self-assemble, so secondary structure and folding dynamics must be encoded in the protein's sequence. Both secondary structure and folding dynamics are important, because proteins must both fold correctly and efficiently.[2] Protein misfolding has been implicated in "protein misfolding" diseases such as Alzheimer's, Parkinson's and prion disease.[3] Treatment of these diseases requires a detailed understanding of the structure and function relationship of proteins. The challenge of finding the relationship between protein sequence and native structure has come to be known as the protein folding problem.

The protein folding problem can be split into three different subproblems: thermodynamics, kinetics and structure prediction.[4] The thermodynamic problem is the question of which interatomic forces in an amino acid sequence dictate the final folded structure. The kinetics problem is the identification of the pathway or the pathways that proteins use to fold quickly to their native structure. The structure prediction problem combines information from the thermodynamic and kinetic studies to design computation methods to predict the native structure from a given amino acid sequence.

In 1972, the Nobel Prize in chemistry was awarded to Christian Anfinsen for his thermodynamic hypothesis of protein folding. Anfinsen discovered that ribonuclease can

be unfolded using denaturant and subsequently refolded into the biologically active structure.[5] From this work, Anfinsen proposed the 'thermodynamic hypothesis' that the native structure of a protein is the thermodynamically stable structure, the structure with the lowest Gibbs free energy. It was concluded that regardless of whether a protein was folded *in vivo* upon biological synthesis, refolded with chaperones or folded *in vitro* it would reach the same native state. Anfinsen's conclusions have enabled researchers to focus their efforts on understanding protein folding *in vitro*.

Because of the large number of degrees of freedom, proteins cannot fold through a random sampling of conformations. For each amino acid in a protein there are 3^N possible conformations. Thus assuming proteins sample a new conformation each nanosecond, the time for a relatively small protein of 100 amino acids to sample every possible conformation would be longer than the age of the universe.[6] However, proteins are known to adopt their native fold much faster, on a millisecond or even microsecond timescale. This contradiction was first pointed out by Cyrus Levinthal in 1969 and has become known as "Levinthal's paradox".[7] Levinthal concluded that protein folding must not be random but rather fold via specific pathways.

A number of protein folding models have been developed in an attempt to resolve Levinthal's paradox. One of the first models, the nucleation-growth model, assumed that the rate-limiting step is a nucleation event, and once nucleation occurs folding proceeds quickly.[8] However, this model was unable to capture the folding of proteins through intermediates. In the diffusion-collision model, microdomains form first and then diffuse and collide, which can lead to adhesion into larger subunits.[9] The framework model and hydrophobic collapse models differ on the importance of short-range contacts. In the

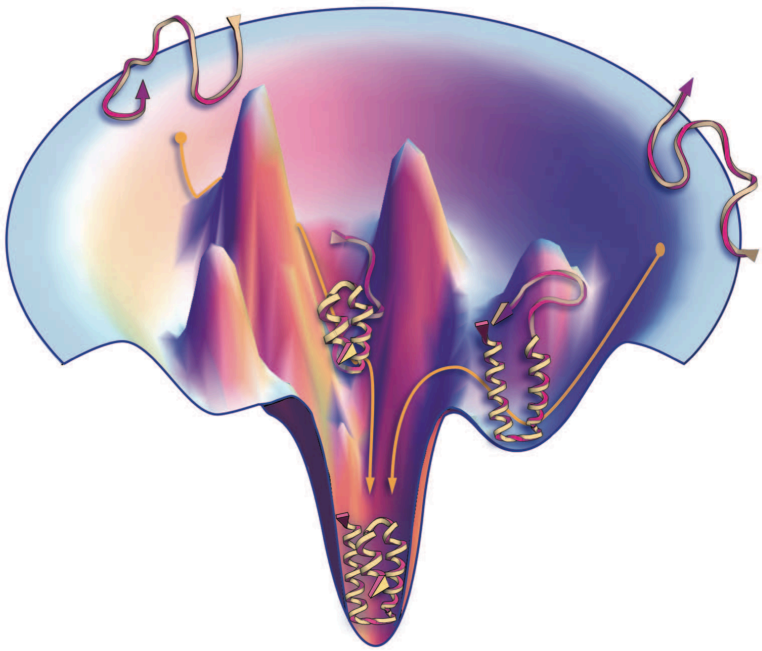


Figure 1-1. Proteins have a funnel-shaped energy landscape with many high energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories.[10] From Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on, *Science 338*, 1042-1046. Reprinted with permission from AAAS.

framework model, short-range interactions drive folding; folding is hierarchical starting with formation of secondary structure.[11] The opposite is true of the hydrophobic collapse model, where greater importance is placed on long-range interactions and folding is initiated by hydrophobic collapse.[12] Finally the jigsaw puzzle model proposed that the folding pathway is not unique; each protein can follow a different route to the native structure.[13]

This view is consistent with the current model of protein folding, the energy landscape.[14-18] The energy landscape can be pictured as a surface in three dimensions. The horizontal axes represent conformational degrees of freedom of the polypeptide chain, and the vertical axis represents the free energy. As the free energy decreases, the polypeptide chain becomes more native with the lowest energy state being the biologically functional native state. The energy landscape of proteins is thought to look like a funnel (Figure 1-1). The energy landscape provides information about the folding kinetics of proteins; smooth funnels lead to fast folding while rough funnels lead to slow folding. Some proteins may get trapped in misfolded states or local energy minima on a rough energy landscape. To understand how proteins fold, we must be able to map this energy landscape.

1.2 Fast Events in Protein Folding

Much effort in protein folding research has gone into studying how proteins go from a folded to unfolded state. Of particular interest are the early kinetic events: which structures form first and on what timescale.[2] Secondary structure formation, chain collapse and diffusion are examples of processes that occur on the submillisecond timescale. This is also a biologically relevant timescale for ultrafast folding proteins, the fastest of which fold in under a microsecond.[19] Evolution of fast folding proteins is one way that nature may balance stability of proteins with the flexibility necessary to function. Many of the fastest folding peptides are subdomains of larger proteins. Another hypothesis is that these small fast folding domains seed the folding of the larger protein. This is a difficult timescale to probe as until recently, a divide existed between what

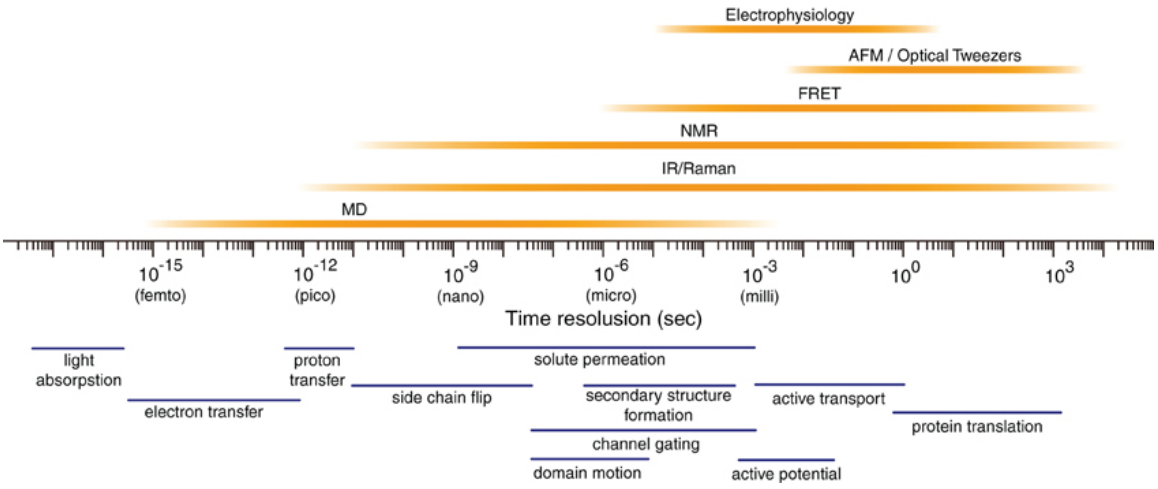


Figure 1-2. Temporal resolution of various biophysical techniques.[20] From Ode, H., Nakashima, M., Kitamura, S., Sugiura, W., and Sato, H. (2012) Molecular dynamics simulation in virus research, *Frontiers in microbiology 3*, 258. This work is licensed under CC BY 4.0. http://creativecommons.org/licenses/by/4.0/

could be modeled and what could be measured experimentally (Figure 1-2).[19] All-atom molecular dynamic simulations were unable to fully fold even a small protein. Experiments were unable to capture the early processes in protein folding.

The most common computational approach to theoretical protein folding is molecular dynamic simulations. One advantage of computational methods is that the experimental parameters and conditions can be adjusted to imitate the conditions of the cell. However, computational results arising from relatively small differences in force fields and potential energy surfaces can make it difficult to compare simulations to each other and with experiment. Simulation of folding events has been limited to small proteins and short timescales because of computational expense. These all-atom methods run into

sampling problems for large proteins. Proteins that fold on a microsecond or longer timescale cannot be observed, because molecular dynamic simulations are limited to the nanosecond to microsecond timescale. The recent development of distributed computing methods and the use of supercomputers have extended this timescale to milliseconds; however, these approaches are limited to a few groups and are not widely accessible.[21, 22] Improvements in theoretical methods allow for the prediction of rates, free energies, mechanism and structure[23]. All-atom molecular dynamic simulations of small fast folding proteins provide structural detail not available experimentally.

What is observed computationally must be validated with experiments. Time resolved spectroscopic methods allow the study of the mechanism of formation of α-helices, β-sheets and the fastest folding proteins.[24] There have been three general classes for rapid initiation of protein folding: photochemical triggering, temperature or pressure jump, and fast mixing methods. These methods are able to probe folding dynamics on a submillisecond timescale. The first fast-folding protein folding study used a photochemical trigger to study photodissociation of carbon monoxide in denatured cytochrome *c*.[25, 26] Turbulent and hydrodynamic focusing mixing experiments have reduced the dead time of mixing experiments to 50-200 μs.[27, 28] Temperature jump experiments use pulsed laser excitation to rapidly initiate protein folding by perturbing the folding and unfolding equilibrium. Time resolved infrared, fluorescence, resonance Raman and circular dichroism can be used to probe the reaction on the nanosecond to millisecond timescale.[19] This approach is more broadly applicable than the other methods, because it can be applied to any process where perturbation produces an enthalpy change and has submicrosecond time resolution. While in many cases there is good agreement

between simulation and experiment, there are still many cases where there is poor agreement[23]. Further combined theoretical and time resolved studies of protein will increase our understanding of protein folding and of how experiment and computation can be used to complement each other.

Small ultrafast folding proteins provide model systems for comparing experiment and theory (Figure 1-3). Many of the fastest folding natural sequences are small helical subdomains of naturally occurring proteins. Alpha helical peptides have a low contact order, interactions between residues that are close in sequence, which has been correlated with fast folding.[29] Because of this, small helical proteins have proven to be good model systems for comparison of experiment and simulation. The first experimental kinetic study of a fast folding protein was of a 21-residue alanine rich peptide.[30] Temperature jumps probed by fluorescence and infrared spectroscopy yielded different results. When probed by infrared the relaxation dynamics were found to be biphasic, a <10 ns phase and 160 ns phase.[30] Fluorescence measurements showed a single <20 ns phase.[31] The nucleation-propagation model was able to explain both experiments; the fast relaxation results from helix propagation, and the slow phase corresponds to crossing the helix nucleation barrier. These results demonstrate that even the folding of small fast folding model proteins cannot be described as a simple transition between folded and unfolded conformations. Experiments and simulations have now been applied to numerous fast folding systems including albumin-binding domain, engrailed homeodomain, B domain of protein A (Appendix I) and villin headpiece subdomain.[19]

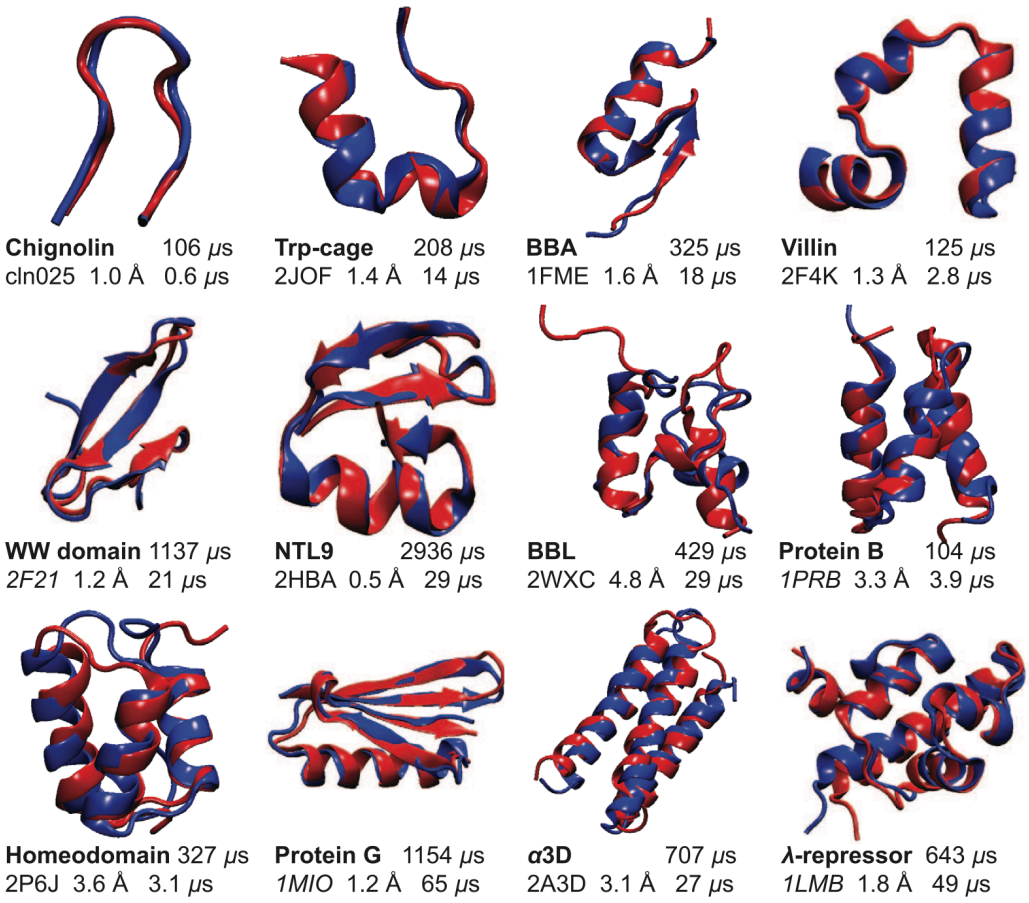


Figure 1-3. Representative structures of the folded states observed in reversible folding simulation of 12 proteins. For each protein the folded structure obtained from simulation (blue) is superimposed on the experimentally determined structure (red), along with the total simulation time, the PDB entry of the experimental structure, the Cα-RMSD (over all residues) between the two structures, and the folding time (obtained as the average lifetime in the unfolded state observed in the simulations).[21] From Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold, *Science 334*, 517-520. Reprinted with permission from AAAS.

How fast can a protein fold? The speed limit for a bimolecular reaction is determined by the rate at which the reactants come together by diffusion. Extending this to protein folding, the rate of folding is limited by the rate of polypeptide collapse and the time required to form a small loop.[32] The protein folding speed limit has been estimated at ~1 μs, a timescale approached by the fastest folding proteins (Figure 1-3). Folding at the speed limit is postulated to occur in the absence of an activation barrier, known as downhill folding.[33] The concept of downhill folding has generated a lot of interest because if downhill folding can be initiated it offers the chance to experimentally follow the accumulation of native structure. Experimental evidence for downhill folding is thought to be the onset of non-exponential, temperature independent kinetics.[34]

1.3 Folding of Beta Proteins

There have been fewer studies of fast folding β-proteins than α-helical proteins. One reason is that isolated β-proteins are prone to aggregation at the high concentrations necessary for NMR experiments to characterize their structures.[35] Another problem is that β-proteins often have low structure content in aqueous solution.[36] Model peptides designed to overcome this are often over-stabilized leading to aggregation.[37] Because of this, there were relatively fewer structures of β-proteins than α-helical proteins available for computer simulations. In addition, β-proteins were not suitable computational models, because folding of β-hairpins is typically ten times slower than the .5 μs of α-helical proteins.[24] Because of the limited studies of β-proteins, the folding mechanism of β-hairpins and their propagation into multi-stranded β-sheets is not fully understood.

The discovery of isolated β-hairpin peptides that do not aggregate has enabled investigation of β-proteins. The first kinetic study of a β-hairpin was of the 16-residue C-terminal hairpin of the protein GB1.[38] The hairpin was found to fold in a few microseconds, an order of magnitude slower than similar sized α-helical peptides. There are two proposed mechanisms of β-hairpin folding which are initiated by hydrophobic and hydrogen bonding interactions.[39] In the turn zipper mechanism folding is initiated by hydrogen bond interactions in the turn. A relatively lower contact order is found in the turn than in the β-sheet of peptide. This mechanism is supported by the observation that low contact order has been correlated with faster folding. In the alternative model, the hydrophobic collapse mechanism is folding is initiated by hydrophobic collapse. In this model a greater importance is placed on long-range interactions.

Infrared spectroscopy is a powerful tool for the analysis of protein structure and folding kinetics. Characteristic IR bands can be assigned to amino-acid side chains and the peptide backbone.[40] The amide I region (1600-1700 cm^{-1}) has been particularly useful for assigning protein secondary structures (Table 1-1). The amide I mode arises from the C=O stretching vibration of the peptide backbone with minor contributions from the out-of-phase CN stretch, CCN deformation and the NH in-plane bend. Using infrared spectroscopy it is possible to independently study the turns and β-sheets of β-proteins. It is possible to further distinguish between solvent exposed carbonyls in the β-sheet and carbonyls hydrogen bonded across the β-sheet. The dipoles of the β-sheets are able to couple, splitting the amide I band. The degree of splitting is dependent on the extent of coupling, which increases with β-sheet length and the number of strands.

Table 1-1: Assignments of amide I band positions to secondary structure in D_2O[40]

	Band Position	
	Average (cm^{-1})	Extremes (cm^{-1})
α-helix	1652	1642-1660
β-sheet	1630	1615-1638
β-sheet	1679	1672-1694
Turns	1671	1653-1691
Disordered	1645	1639-1654

Fourier transform infrared spectroscopy (FTIR) was used in the research presented in this dissertation to measure equilibrium temperature-dependent absorption of β-proteins. From the temperature dependent data a difference spectra was constructed to identify the specific secondary structure elements of the protein that change with temperature. Equilibrium FTIR measurements were complemented by temperature dependent fluorescence spectroscopy and circular dichroism. The spectral assignments and temperature dependent absorbance changes obtained from these measurements provided a control for kinetic experiments. β-protein kinetics were measured by laser induced temperature jumps coupled with infrared or fluorescence spectroscopy. Results from these experiments will provide more information on the folding mechanisms of β-proteins.

1.4 Dissertation Outline

Specific Aim 1: to characterize the folding mechanism and dynamics of a β-hairpin and WW domain *Hypothesis: folding of β-sheet structures is heterogeneous*. We have taken a hierarchal approach to studying β-sheet propagation, starting from the smallest β-sheet structural motif, the β-hairpin, and building up to the WW domain, which incorporates 2 β-hairpins. A combination of Fourier transform infrared spectroscopy and laser induced temperature jumps coupled with infrared and fluorescence spectroscopy were used to probe changes in the peptide backbone. We find that even for the fastest folding linear β-hairpin, CLN025, folding is heterogeneous and cannot be described by a simple two state model (Chapter 2). FBP28 WW domain exhibited multiple relaxation lifetimes covering the nanosecond to millisecond timescale (Chapter 3). These events have not been reported in previous experimental or computational studies.

Specific Aim 2: to study the effect of insertion of a fast folding hairpin into larger β-sheet structures *Hypothesis: insertion of a fast folding hairpin into larger peptides speeds protein folding*. What is the role of the hairpin in β-sheet formation? In Specific Aim 1 we observe that CLN025 folds on the nanosecond timescale, whereas FBP28 folds on the microsecond timescale. We have systematically replaced each of the loops of FBP28 with CLN025 (Chapter 4). Optimization of the first turn of another WW domain, Pin1, resulted in an increase in the thermal stability and folding kinetics (Chapter 5). Using the methods described in Specific Aim 1 we find that formation of WW domains is hierarchical, initiated in the first hairpin. These studies demonstrate that extended β-sheet

structures can be seeded from a hairpin and further than the hairpin stability affects the speed of β-sheet propagation.

Specific Aim 3: to test the proposed model of hairpin formation and insertion into a membrane *Hypothesis: peptide binding to the membrane surface neutralizes peptide charge allowing the hairpin to fold, insert and disrupt the membrane*. Does hairpin folding occur on the surface of the membrane or after insertion? SVS-1, an 18-residue anti-cancer peptide that is designed to be disordered in solution, but fold into a β-hairpin at an anionic bilayer surface, was used as a model system for protein folding on the membrane (Chapter 6). A combination of equilibrium circular dichroism and fluorescence spectroscopy was used to monitor peptide secondary structure changes and insertion into the hairpin. We observe that the peptide folds on the surface of charged lipids, but insert only into membranes in the fluid phase. Insight from these studies can be applied towards development of drugs for the treatment of cancer.

Chapter 2: Raising the Speed Limit for β-Hairpin Formation[41]

2.1 Abstract

Understanding the folding of the β-hairpin is a crucial step in studying how β-rich proteins fold. We have studied CLN025, an optimized ten residue synthetic peptide, which adopts a compact, well-structured β-hairpin conformation. Formation of the component β-sheet and β-turn structures of CLN025 was probed independently using a combination of equilibrium Fourier transform infrared (FTIR) spectroscopy and laser induced temperature-jump (T-jump) coupled with time resolved infrared and fluorescence spectroscopies. We find that CLN025 is an ultrafast folder due to the small free energy barrier to folding and that it exceeds the predicted speed limit for b-hairpin formation by an order of magnitude. We also find that the folding mechanism cannot be described by a simple two-state model, but rather is a heterogeneous process involving two independent parallel processes. Formation of stabilizing cross-strand hydrophobic interactions and turn alignment occur competitively, with relaxation lifetimes of 82 ± 10 ns and 124 ± 10 ns, respectively, at the highest probed temperature. The ultrafast and heterogeneous folding kinetics observed for CLN025 provide evidence for folding on a nearly barrierless free energy landscape, and recalibrate the speed limit for the formation of a β-hairpin.

2.2 Introduction

Small single-domain peptides that fold quickly provide experimental models that are useful for studying the fundamental principles of protein folding.[2, 42] Peptide models of both α-helix and β-sheet have provided important insight about the stability and dynamics of these secondary structures. Extensive studies of model α-helical peptides show that the rate of formation of the α-helix is fast, with a characteristic lifetime on the sub-microsecond timescale.[43-45] Similar studies on model β-hairpin peptides reveal slower folding rates that span a wide range.[46-48] These differences can be understood in terms of the speed limit for the formation of each secondary structural element, a concept that emerges from the free energy landscape theory of protein folding. The speed limit for folding is a consequence of folding on a smooth, funneled free energy landscape with a minimal free energy barrier to folding.[33] Eaton introduced the important concept of a speed limit for β-hairpin formation based on the minimum chain collapse time to form stabilizing cross-strand interactions (between residues that are far apart in sequence), whereas the rate of helix formation is limited by the helix nucleation time, determined by local interactions. Therefore, the speed limit for β-hairpin formation is estimated to be $\sim 10^6$ s^{-1}, approximately tenfold slower than the speed limit of α-helix formation.[24] These ideas have been difficult to test experimentally. Proteins often exhibit two-state folding behavior, meaning they fold on an energy landscape with a single dominant energy barrier. Two-state folders exhibit simple single-exponential kinetics, representing population flow over the barrier, without providing any microscopic insight into the reaction progress. The same rate is observed regardless of the structural probe employed. But folding in the absence of a barrier should occur at the speed limit, and in this case it

should be possible to observe the progressive development of the folded structure directly, revealing the complexity and heterogeneity of the process. Thus, a defining characteristic of folding at or near the speed limit should be non-two-state folding behavior.

The β-hairpin is the simplest β-sheet structural motif, and it is often found in larger β-sheets in globular proteins. It is possible to form soluble and stable β-hairpins from isolated short peptides,[49] supporting the idea that they may act as nucleation sites for folding more complex β-sheet structures. Small fast-folding model β-hairpins are ideal candidates for protein folding studies of β-proteins, because the folding dynamics of these systems occur on a timescale where experiment and all-atom molecular dynamic simulations overlap with each other. Previously, we have measured relaxation kinetics of model β-hairpins using laser induced temperature jump (T-jump) and time-resolved infrared (IR) and fluorescence spectroscopy.[50, 51] Similar methods have been used by a number of groups to study the folding of β-hairpins. Using fluorescence T-jump, Yang et al. reported wavelength-dependent relaxation lifetimes of trpzip2 on the microsecond timescale.[52] Xu et al. studied folding kinetics of a 15-residue β-hairpin that folds within 800 ns at 300 K.[53] Our results for a series of cyclic β-hairpin analogues of Gramidicin S exhibited relaxation lifetimes ($\tau \sim 100$ ns) independent of chain length, much faster than those reported for linear β-hairpins.[50] These results suggest that the folding rates of β-hairpins are not uniform, but rather are dependent on many factors, particularly the specific type of turn structure and the rate of chain collapse.

The CLN025 β-hairpin (YYDPETGTWY, Figure 2-1) is a good model system for further investigation of β-hairpin folding dynamics, because its stability and small size

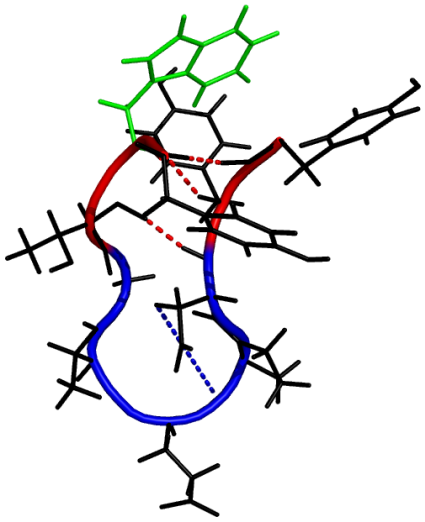


Figure 2-1. Cartoon of the CLN025 crystal structure[54] with tertiary amide linkage in the turn (blue) highlighted with blue dashed line, intermolecular hydrogen bonded carbonyl groups in the β-sheet (red) highlighted by red dashed lines and fluorescent tryptophan highlighted green.

are ideal for ultrafast folding experiments. CLN025 is a member of the designed Chignolin family of peptides. The parent Chignolin is a ten residue β-hairpin derived from the sixteen residue β-hairpin G-peptide.[55] The internal eight residues of Chignolin were statistically optimized to form a stable β-hairpin using a databank of over 10,000 peptide sequences.[56] The terminal residues of Chignolin were further optimized to design CLN025.[54] Computational studies on Chignolin and CLN025 have predicted folding lifetimes of hundreds of ns,[21, 57] the same timescale as the fastest previously reported β-hairpins. Computational studies have not agreed on the folding mechanism of Chignolin with one study reporting the turn forming first[58] and another reporting hydrophobic

collapse as the first step.[59] Lindorff-Larsen *et al*. simulations of CLN025 predict hairpin formation along multiple folding pathways.[21] These predictions for Chignolin and CLN025 mirror the general state of understanding of β-hairpin formation. Two different models have been proposed, one initiated by hydrophobic and the other by hydrogen bonding interactions.[39] In the first model, the turn zipper mechanism postulated by Muñoz *et al*., folding is initiated in the turn region and zipped from the turn to the end.[30, 43, 60] Dinner et al. proposed an alternative model, the hydrophobic collapse mechanism, which predicts hydrophobic collapse prior to formation of hydrogen bonds and the hairpin.[46-48]

We have studied the dynamics of β-turn formation and hydrogen bond propagation in CLN025 using T-jump, time-resolved infrared spectroscopy. Pulsed laser excitation was used to rapidly initiate a shift in the folding equilibrium. The relaxation dynamics of CLN025 were measured by independently probing the amide I components assigned to turn formation and β-sheet formation. The IR studies were complemented by fluorescence T-jump experiments that probe tryptophan (Trp) side chain dynamics.[19] We find that CLN025 folds much faster than any previously studied linear b-hairpin, consistent with the timescale predicted by computations.[21, 57] We find that CLN025 has a small free energy barrier to folding and the observed folding rate exceeds the predicted speed limit, likely because chain collapse is not rate limiting in this case. Instead of a dominant pathway involving either turn formation or hydrophobic collapse as a rate limiting step, we observe heterogeneous folding rates that depend on the specific structural probe. Using multiple spectroscopic probes we observe formation of stabilizing cross-strand hydrophobic interactions and turn formation occurring in parallel but on

slightly different timescales. The ultrafast and heterogeneous folding kinetics provide evidence for folding at the speed limit on a nearly barrierless free energy landscape.

2.3 Experimental Section

Protein Synthesis and Purification. The CLN025 peptide, YYDPETGTWY,[54] was synthesized via standard 9-fluornylmethoxycarbonyl (Fmoc) based solid-phase chemistry. Fmoc-Tyr(tBu)-PEG-PS resin was used to form a free C-terminus. The peptide was purified by reverse-phase high-pressure chromatography. HCl was used as the counter-ion to reduce the residual TFA concentration since TFA can interfere with Amide-I IR measurements at 1672 cm^{-1}. The identity of the peptide was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. The peptide was lyophilized and dissolved in D_2O to allow deuterium-hydrogen exchange of the amide protons. The peptide was lyophilized a second time and resuspended in a filtered D_2O buffer with 20 mM potassium phosphate and 150 mM sodium chloride at pH 7.0. Sample concentrations of 0.5-6 mM were prepared for both IR and fluorescence experiments.

FTIR Spectroscopy. The equilibrium melting behavior was monitored on a Varian Excalibur 3100 FTIR spectrometer (Varian Inc., Palo Alto, USA) using a temperature controlled IR cell. The IR cell consists of two CaF_2 windows stacked and separated by a 100 μm Teflon spacer split into two compartments, a sample and a reference. The same cells are used for equilibrium FTIR and T-jump experiments. No aggregation was observed (characteristic amide I' bands indicative of aggregation were not observed) at reported concentrations. All spectra shown at a specific temperature are constructed by

subtracting the spectrum of reference buffer solution without protein from sample solution with protein. The temperature-dependent difference spectra were then generated by subtracting the spectrum at 5 °C from the spectra at higher temperatures. The temperature-dependent difference spectra were baseline corrected in the amide I' region from 1500-1750 cm^{-1} using a linear baseline to account for the baseline differences caused by slight differences in H_2O concentration. The second derivative was performed in IGOR PRO (WaveMetrics, Lake Oswego, OR).

Time Resolved Temperature Jump (T-jump) Relaxation Measurements. The T-jump apparatus has been described previously.[30] Pulsed laser excitation is used to rapidly perturb the folding equilibrium on a timescale faster than the molecular dynamics of interest. Time resolved infrared or fluorescence is then used to probe the reaction. A Q-switched DCR-4 Nd:YAG laser (Spectra Physics, Mountainview, CA) fundamental at 1064 nm is Raman shifted (one stokes shift in 200 psi H_2 gas) to produce a 10 ns pulse at 2 μm. The magnitude of the T-jump is calculated using the change in reference absorbance with temperature. The T-jump reference is taken from the 6 mM sample at a frequency, 1680 cm^{-1}, that is temperature independent for the peptide based on the equilibrium FTIR studies. Absorbance changes at the reference wavelength are only due to changes in D_2O absorbance, which is used as an internal thermometer.[30]

The change in signal induced by the T-jump is probed in real time by a continuous laser with a wavelength either in the amide I' band of the IR or at 285 nm to excite Trp fluorescence. The mid-IR probe beam is generated by a continuous wave quantum cascade laser (Daylight Solutions Inc., San Diego, CA) with a tunable output range of 1535-1695 cm^{-1}. The transient transmission of the probe beam through the sample is

measured using a fast, 100 MHz, photovoltaic MCT IR detector/preamplifier (Kolmar Technologies, Newburyport, MA). Transient signals are digitized and signal averaged (5000 shots) using a Tektronics digitizer (7612D, Beaverton, OR).

The fluorescence excited at 285 nm and collected from 320-370 nm is sensitive to changes in the tryptophan residue near the C-terminus of the peptide. A Verdi V12 DPSS High-Power continuous wave laser (Coherent, Santa Clara, CA) is used to pump a Mira 900 Ti:Saph laser (Coherent, Santa Clara, CA) which produces a quasi-continuous beam at 855 nm. The beam is then passed through a second and third harmonic generator (Coherent, Santa Clara, CA) to produce the fluorescence probe beam at 285 nm. The back-emitted fluorescence light induced by the 285 nm pump laser is measured through a 355 ± 15 nm band pass filter using a Hamamatsu R7518 photomultiplier tube (Hamamatsu Photonics K.K., Hamamatsu City, Japan), digitized and signal averaged (10,000 shots) using a Tektronics digitizer (7612D, Beaverton, OR). A 400 μM tryptophan (Trp) solution in D_2O is used as a reference to measure the magnitude of the temperature jump and determine the temperature dependence of the tryptophan signal. Instrument control and data collection are controlled using a LabVIEW computer program.

Analysis of Kinetics Data. A reference is collected for each sample measurement at each temperature. The relaxation lifetime of the reference signal is the instrument response of the sample. In order to minimize detector artifacts, the reference is scaled prior to subtraction from the sample. This process leaves a small residual signal from the instrument response. Due to the fast timescale, accurate determination of the peptide relaxation kinetics requires a deconvolution of the instrument response function from the

observed kinetics. Standard deconvolution methods were unable to accurately deconvolve the instrument response from the relaxation kinetics because of fast timescale artifacts in the data. Instead the instrument response and peptide relaxation kinetics were fit to a biexponential function given as:

$$y = A_0 + A_1 \exp\left(\frac{-(x - x_0)}{\tau_1}\right) + A_2 \exp\left(\frac{-(x - x_0)}{\tau_2}\right) \qquad (2\text{-}1)$$

where A_0 is an offset, A_1 and A_2 are preexponential factors, τ_1 is the instrument response, τ_2 is the relaxation lifetime of the sample and x is the time. The data was fit to a biexponential function and the first relaxation lifetime, corresponding to the instrument response, was averaged over all temperatures to get a global instrument response value. In a second set of fits to the biexponential function the first relaxation lifetime was fixed to the global instrument response value, 20 ns for IR T-jumps and 50 ns for fluorescence T-jumps, both in good agreement to the previously reported instrument response time of ~20 ns.[61] The reported kinetics are from the second relaxation lifetime in the biexponential function. All data was fit from 10^{-8} to $2x10^{-6}$ s. The data analysis was performed in IGOR PRO (WaveMetrics, Lake Oswego, OR).

Results

2.4 Equilibrium FTIR Studies. The temperature induced unfolding of CLN025 was studied over the range from 5 to 85 °C in 5 °C intervals using FTIR spectroscopy to monitor the amide I' region. The amide I' spectral region (amide I region of peptides in D_2O), composed of mostly C=O stretching vibrations of the polypeptide backbone carbonyls, is an established indicator of secondary structure.[62-64] The absorption spectra of the amide I' region as a function of temperature are shown in Figure 2-2 A. This

relatively broad band contains contributions from the entire polypeptide backbone, which in the case of the β-hairpin is made up of contributions from the β-sheet, the β-turn, or any random coil regions. The changes with temperature are highlighted in the difference spectra for each peptide, in Figure 2-2 B. The difference spectra are generated by subtracting the lowest temperature spectrum from each absorbance spectrum at higher temperature. Negative peaks correspond to specific structures or interactions present in the folded state; the strongest of these are the peaks at 1614 and 1629 cm^{-1}. Positive peaks correspond to new interactions with solvent in the unfolded state; the characteristic feature due to the disordered polypeptide is a relatively broad peak, located at 1660 cm^{-1}.[30, 43, 65] The broad nature of these peaks may conceal some of the bands due to ordered structure. Spectral interpretation was aided by taking the second derivative of the FTIR spectra (Figure 2-2 C) to uncover otherwise indistinguishable features.

The analysis revealed five components of the amide I' band at 20 °C, centered near 1614, 1629, 1649, 1658 and 1678 cm^{-1} (Figure 2-2 C). The intensity of these features decreases with increasing temperature, meaning these features are all associated with the folded state. The Keiderling group identified a characteristic peak at ~1611 cm^{-1} in β-hairpins with a proline in the turn, which they attributed to a tertiary amide linkage between the proline and the side chain of a neighboring amino acid.[66] Similarly, we assign the peak at 1614 cm^{-1} to a tertiary amide linkage between the backbone of glutamic acid and the side chain of aspartic acid in the turn (Figure 2-1).[56] Two peaks are observed for the β-sheet, corresponding to the inter- and intramolecular hydrogen bonded carbonyl groups.[50] The peak at 1629 cm^{-1} is assigned to the inward directed carbonyl groups (Figure 2-1) and the peak at 1678 cm^{-1} is assigned to the solvent exposed carbonyl

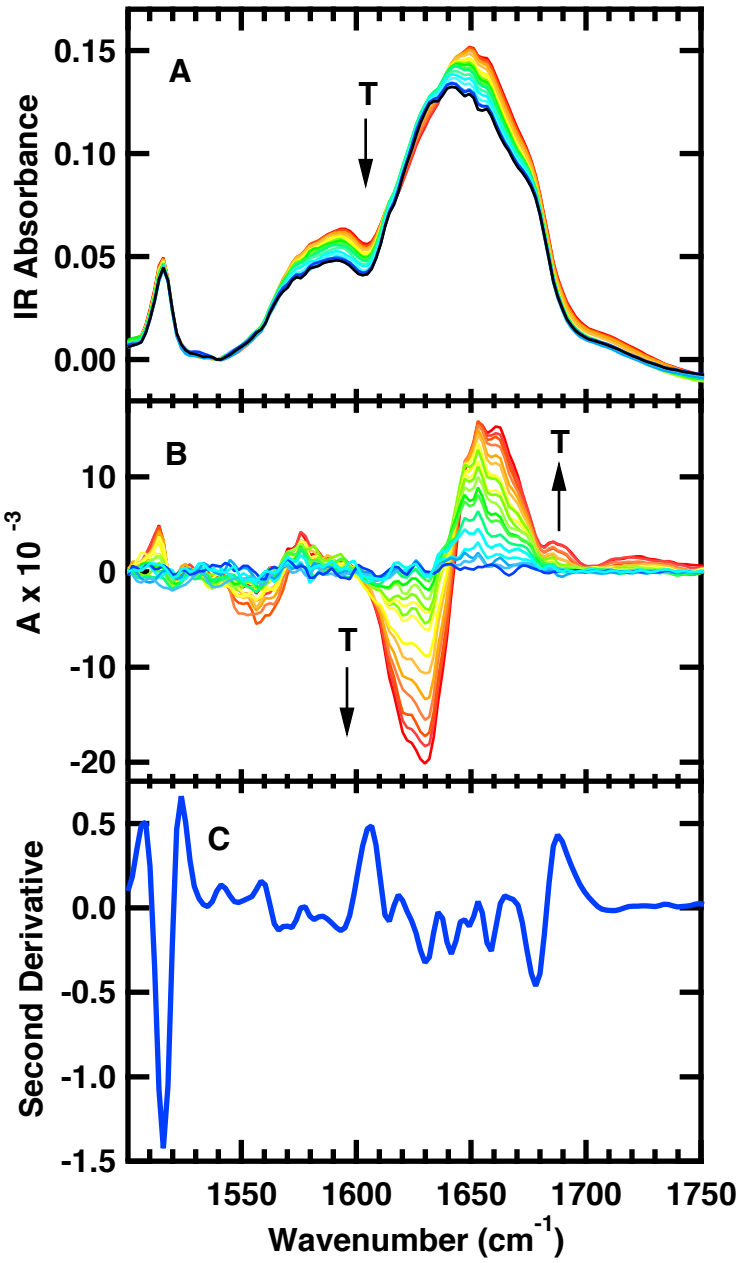


Figure 2-2. Temperature dependent FTIR spectra of 6 mM CLN025, in 20 mM potassium phosphate buffer (pH 7). (A) Absorbance spectra in the Amide I' region; the temperatures of the individual traces varies from 5 to 85°C from bottom to top in 5 °C intervals. (B) Baseline corrected difference spectra obtained by subtracting the spectrum at 5°C from the spectra at higher temperatures. (C) Second-derivative spectrum at 5 °C.

groups. These assignments are supported by several pieces of evidence.

The FTIR spectral amide I' absorbance maximum at 1629 cm^{-1} agrees with the well-documented indicator of cross-strand β-sheet interaction at 1634 cm^{-1}.[66] Most of the amide carbonyl groups are located in the sheet portion of the hairpin: there are six sheet carbonyls, four of which participate in intramolecular hydrogen bonds and two in intermolecular hydrogen bonds to water, as opposed to 3 carbonyls in the turn, involved in one tertiary amide, one intramolecular and one intermolecular hydrogen bond to water, respectively. The relative intensities of the sheet and turn bands support these assignments if the individual C=O groups are assumed to act as local oscillators and have similar oscillator strengths. The high frequency bands at 1649, 1658 and 1678 cm^{-1} overlap with the broad positive absorbance centered at 1660 cm^{-1} that grows in with increasing temperature in the difference spectra. This broad random coil peak is composed of carbonyl groups hydrogen bonded to water, so solvent exposed carbonyls in the folded structure are also expected to have similar frequency. The additional two peaks at 1649 and 1658 cm^{-1} cannot be positively identified, however, they most likely correspond to β-turn carbonyls with different degrees of solvent exposure or interaction.[66]

The melting curve derived from the temperature dependent IR absorbance at 1629 cm^{-1} is shown in Figure 2-3. To test whether the peptide exhibits a simple two-state behavior, the data is fit to an apparent two-state equilibrium model:

$$A = \frac{A_i + A_f \cdot 10^{2(T-T_m)/\Delta T}}{1+10^{2(T-T_m)/\Delta T}} \qquad (2\text{-}2)$$

where A_i and A_f are the extrapolated absorbance values at the two endpoints of the transition, T_m is the transition midpoint and ΔT represents the overall temperature range

of the transition. The T_m and ΔT values from the two-state model fit are listed in Table 2-1. The melting temperature of the β-sheet, 70.9±1.8 °C, agrees well with the previously reported melting temperature obtained by circular dichroism, 69.9 °C.[54] The melt curve derived from other wavelengths agreed with the melting temperature within error of the fit. The relatively sharp transition (small ΔT) reveals that the melt is cooperative.[50] Apparent thermodynamic parameters (assuming a two-state model) derived from a van't Hoff analysis (inset of Figure 2-3) are also summarized in Table 2-1. A van't Hoff plot of the log of the observed equilibrium constant versus 1/T should be linear. The van't Hoff analysis reveals a pre-transition below 35 °C, showing that hairpin folding is not a simple two-state process. The non-linear behavior cannot be explained by a heat capacity change upon unfolding. It is unlikely that such a small structure would have a large difference in heat capacity between the folded and unfolded states, however, if it did it would have a parabolic dependence on temperature, which does not agree with the observed line shape. Previous studies on Chignolin, which is the parent of CLN025, reported a $\Delta C_p \approx 0$.[56]

Thermodynamic parameters were obtained by treating the pre-transition, below 35 °C, and transition, above 35 °C, as two independent two-state processes. The enthalpy gain and entropic cost for the higher temperature transition are nearly double those of the lower temperature one. Both theoretical and experimental studies have shown that vibrational spectroscopy can be sensitive to solvent interaction with the peptide backbone.[67] Therefore we assign the low temperature transition, with its relatively small enthalpic difference to a subtle global change in solvent interaction. In contrast, the higher temperature transition with a larger enthalpic change is clearly due to the hairpin folding.

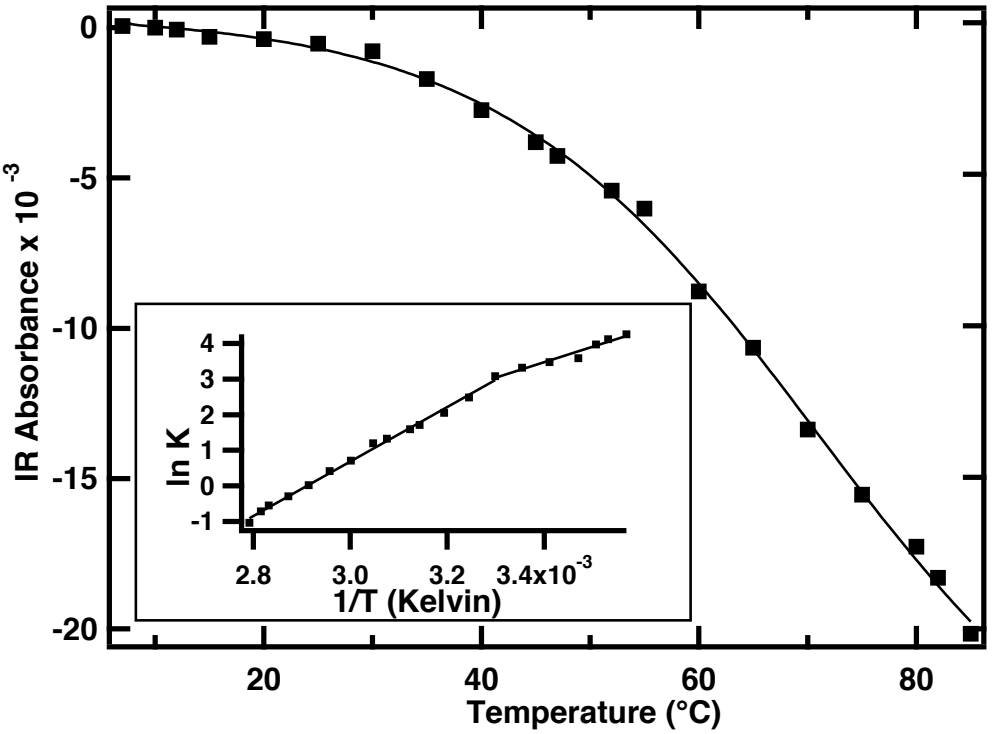


Figure 2-3. FTIR melt curves for the CLN025 peptide obtained by plotting the change in the baseline corrected IR difference spectra intensity at the peak maximum of the β-sheet band at 1629 cm^{-1} versus temperature. The solid line is fit to a sigmoid (Eq. 2-1). Inset: Van't Hoff analysis using equilibrium constants generated from the melt curve.

Table 2-1: CLN025 thermodynamic parameters of folding

	T_m (°C)	ΔT (°C)	ΔH_f (kJ/mol)*	ΔS_f (J/mol•K)*
CLN025 < 35 °C	70.9 ± 1.8	67.1 ± 4.1	-37.0 ± 1.5	-96.8 ± 5.1
CLN025 ≥ 35 °C	70.9 ± 1.8	67.1 ± 4.1	-62.8 ± 1.3	-182.8 ± 3.9

*Thermodynamic parameters are derived from a two-state model and a van't Hoff analysis

2.5 Temperature-Jump Relaxation Kinetics. The probe dependent relaxation kinetics of the folding/unfolding transition following a laser induced temperature jump were probed using time-resolved infrared spectroscopy of the amide I' frequency for the tertiary amide of the turn (1619 cm^{-1}) and the β-sheet (1629 cm^{-1}). Jumps in the turn region were performed slightly off the peak center (1614 cm^{-1}) to maximize the signal. Kinetic experiments examined the dependence of the relaxation rates on the final temperature following a T-jump. The magnitude of the T-jump was kept constant while varying the final temperature with all final temperatures below the melting transition. Figure 2-4 shows the results for a T-jump from 40 to 55 °C probed by IR absorbance. The transients were fit to a double exponential. The first relaxation lifetime for all jumps corresponds to the solvent response, which occurs within the instrument response time, and the second lifetime, which decreases with increasing temperature, corresponds to the relaxation lifetime of the specific peptide structure being probed. The relaxation kinetics observed for the peptides are reported in Table 2-2. At low final jump temperatures there is little difference between the turn and β-sheet kinetics. As the final jump temperatures increase the turn relaxation lifetimes, probed at 1619 cm^{-1}, stabilize at 124 ± 10 ns. The β-sheet lifetimes, probed at 1629 cm^{-1}, continue to decrease throughout the entire series of jumps with a relaxation lifetime at the final jump temperature of 82 ± 10 ns, significantly faster than the relaxation lifetimes observed in the turn. The difference in the relaxation lifetimes observed in the turn and the sheet is well outside the error in the measurements. Because the final temperature of all jumps is at least 15 °C below the melting temperature, the folding rate will dominate the observed relaxation kinetics.

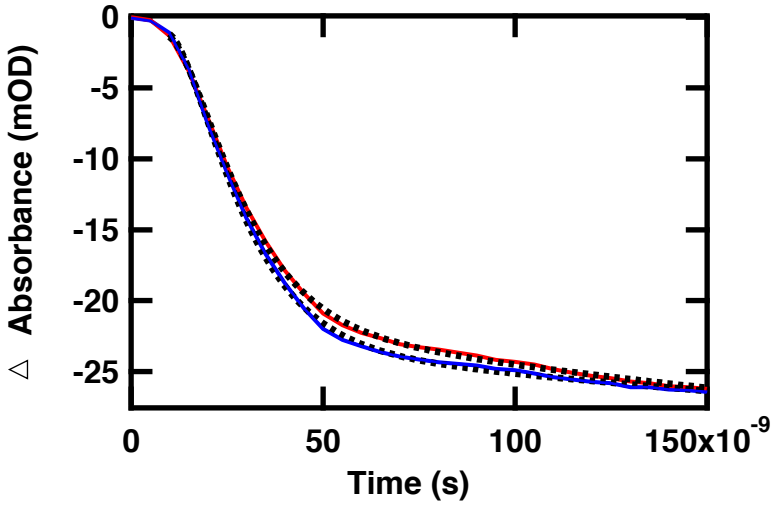


Figure 2-4. Representative T-jump relaxation kinetics monitored in the amide I' spectral region at 1619 cm^{-1} (blue) and 1629 cm^{-1} (red) following a T-jump from 40 to 55 °C. A double exponential fit (dashed line) is overlaid on each kinetics trace. The amplitudes are normalized to the 1629 cm^{-1} amplitude at long time (4 μs) to aid comparison of the relaxation times.

A second set of kinetics experiments with a series of similar magnitude jumps were performed using time-resolved fluorescence. There is a tryptophan located near the terminus of the hairpin, so the tryptophan fluorescence intensity is sensitive to changes in the β-sheet structure and hydrophobic packing of the core of the hairpin (Figure 2-1). Figure 2-5 shows the results for a temperature jump from 15 to 25 °C probed by fluorescence emission. Similar to the infrared temperature jumps, we observe an instrument limited response component (due to the intrinsic temperature dependence of the Trp fluorescence quantum efficiency) in addition to the fluorescence relaxation

Table 2-2: Relaxation Kinetics for CLN025

IR T-jump $\Delta T = T_i - T_f$ (°C)	**τ_r (ns)** ν = 1619 cm^{-1}	**τ_r (ns)** ν = 1629 cm^{-1}	**Fluorescence T-jump** $\Delta T = T_i - T_f$ (°C)	**τ_r (ns)** λ_{cm} = 350 nm
5 - 20	256 ± 34	257 ± 30	10 - 20	261 ± 103
10 - 25	209 ± 23	173 ± 16	15 - 25	196 ± 44
15 - 30	158 ± 53	193 ± 44	20 - 30	131 ± 9
20 - 35	110 ± 8	107 ± 7	25 - 35	103 ± 8
25 - 40	111 ± 8	100 ± 8	30 - 40	95 ± 20
30 - 45	129 ± 14	98 ± 8	35 - 45	95 ± 92
35 - 50	124 ± 10	94 ± 8	40 - 50	91 ± 67
40 - 55	124 ± 10	82 ± 10		

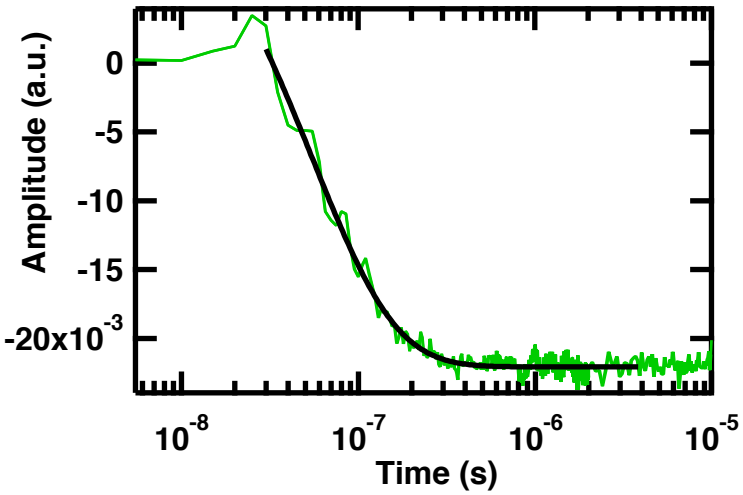


Figure 2-5. Representative fluorescence T-jump relaxation kinetics excited at 285 nm and monitored at 350 nm following a jump from 15 to 25 °C. A double exponential fit (dashed line) is overlaid on the kinetic trace.

kinetics (Table 2-2). As expected, the fluorescence probed relaxation lifetimes are similar to the IR probed relaxation lifetimes of the β-sheet.

Both fluorescence and infrared probed jumps had a final jump temperature below the melting transition, so the observed kinetic rates are dominated by the folding rate. If the β-hairpin followed two-state folding kinetics, an Arrhenius plot (Figure 2-6) of the natural log of the measured rate constants versus 1/T should be linear for the case where there is no heat capacity change upon folding. The Arrhenius plot shows three distinct regions separated by temperature and probed region of the peptide. The kinetics observed for the fluorescence or the IR probed at 1629 cm^{-1} is sensitive to changes in the sheet, whereas the kinetics probed by IR at 1619 cm^{-1} is sensitive to changes in the turn. The break in temperature occurs at 35 °C, the same break that was observed in the equilibrium van't Hoff analysis. The observed activation energy in this region is the same for all fluorescence and IR probed kinetics. This supports the conclusion from the van't Hoff analysis that there is a global change that affects both the turn and β-sheet at lower temperatures, since both the turn and β-sheet experience the same activation energy. Above 35 °C the Arrhenius plot shows different behavior depending on the region of the peptide that is probed. The β-sheet region exhibits faster kinetic rates and a small apparent activation energy. The turn region shows a slower kinetic rate and an apparent activation energy near zero. This analysis does not account for the temperature dependence of the viscosity of the solution.

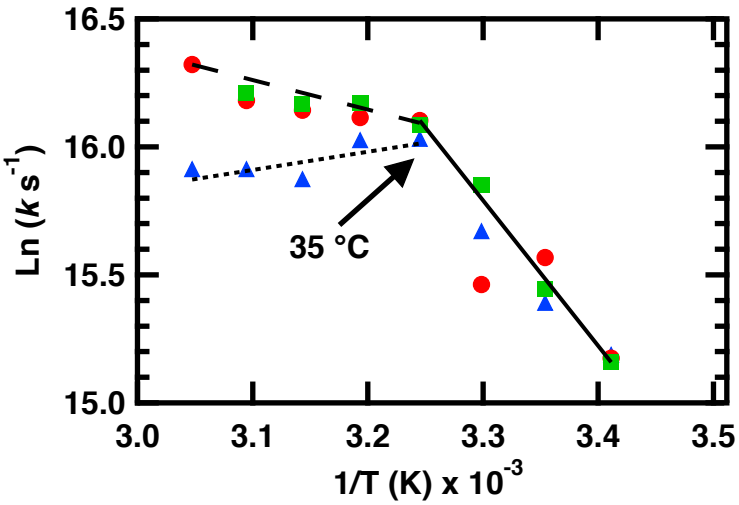


Figure 2-6. Arrhenius plot of relaxation kinetics below T_m. The values of T used for the (1/T) axis are the final temperatures reached during the jump (all below T_m). *k* is the relaxation rate obtained from a fit (see text) of the T-jump transient fluorescence (■) or infrared monitored at 1629 cm^{-1} (●) and 1619 cm^{-1} (▲). Lines are the result of fitting all data (■, ●, ▲, solid line), infrared 1629 cm^{-1} and fluorescence (■, ●, dashed line), and infrared 1619 cm^{-1} (▲, dotted line).

2.6 Discussion

Simple peptide models are valuable as a basis for understanding the fundamental processes of protein folding that lead to the complex architecture of globular proteins. β-hairpins are the simplest structural motif of a β-protein, and yet their folding mechanism is not fully understood. Previous protein folding studies predict β-hairpin formation to follow a turn zipper mechanism or a hydrophobic collapse mechanism and to fold with a speed limit that is significantly slower than that of a-helix.[30, 39, 43, 60] We have studied these

issues for the CLN025 peptide, an optimized β-hairpin structure with a unique hydrophobic cluster at its terminus.

A van't Hoff analysis is used to gain insight into the nature of the thermally induced transition. Over the temperature range from 5 to 85 °C we observe a global pre-transition at low temperature. The enthalpy gain and entropic cost at higher temperature are nearly double those at lower temperature. The apparent thermodynamic parameters are similar to those found previously for a 10-residue cyclic hairpin ($\Delta H° = -48.3$ kJ mol^{-1}, $\Delta S° = -140.4$ J mol^{-1} K^{-1}).[50] Making the assumption that the thermodynamic parameters at higher temperature (but still well below T_m) are the thermodynamic parameters of folding, the entropy loss of folding of CLN025 is thirty percent greater than that of a cyclic peptide. The cyclization study showed that side chain motions dominate the entropy change in the constrained peptide. CLN025 likely has a similar contribution to ΔS from side chain motions, however, we expect that there is an additional small entropic contribution from the peptide backbone motion since it is not fully constrained.

We used IR probes to study the relaxation kinetics of the turn (1619 cm^{-1}) and the sheet (1629 cm^{-1}), and a fluorescence probe to study relaxation kinetics of the tryptophan side chain. Previously, this multiple-probe approach was successfully used to probe the inter- and intramolecular hydrogen bonds in the villin headpiece subdomain.[68] We see a break between low and high temperature regions, at the same temperature as that observed in the van't Hoff analysis. At low temperatures we observe the same relaxation kinetics regardless of time-resolved probe, supporting the conclusion that this process is a global solvent rearrangement. At temperatures above the discontinuity we observe two sets of relaxation kinetics, depending on the region of the peptide that is probed. The folding

kinetics cannot be fit with a simple two-state model. The relaxation lifetimes suggest that the β-sheet and turn form in parallel with slightly different rates, the β-sheet having a faster relaxation lifetime. Studies on type II' β-turns have shown that the rate is dominated by the folding of the turn structure.[50] Our results support this analysis, and extend the observation to non-type II' turns. The turn is found to have a near zero apparent activation energy, consistent with the ultrafast folding rate. Any remaining barrier to turn formation must be small and mostly entropic, a consequence of the search to find stabilizing contacts and the correct alignment of the turn.

The T-jump relaxation kinetics observed for CLN025 are three times faster than the fastest relaxation kinetics observed for other linear β-hairpins.[51, 53] The observed relaxation rates are two orders of magnitude faster than those observed for the β-hairpin of protein G.[38] The β-hairpin of protein G provided a model for the derivation of CLN025.[54-56] The statistical optimization of the β-hairpin sequence appears to have greatly accelerated the folding of the peptide. The relaxation kinetics of CLN025 are on the same order as the relaxation kinetics of cyclic peptides.[50] In a cyclic β-hairpin the ends of the structure are "pinned" together (covalently bonded), restricting the accessible conformations in the unfolded state. This conformational restriction eliminates chain collapse as the rate-limiting step. The interactions between aromatic and hydrophobic groups at the termini of CLN025 may play a similar role and serve to severely restrict conformation space and help to drive folding by rapidly forming a collapsed structure. We postulate that there are residual interactions between these hydrophobic and pi-stacking aromatics in the unfolded state that stabilize collapsed conformations in which the termini are pinned together, analogous to what is observed in cyclic hairpins, leading

to similar folding dynamics. This model also explains why the folding rate exceeds the speed limit, because chain collapse is not the rate limiting step.

The low activation barrier to β-hairpin formation allows us to observe subtle differences in the free energy landscape, depending upon how the reaction is probed. An Arrhenius plot of the different probe kinetics shows three distinct types of behavior. The Arrhenius plot shows that there is a small enthalpic barrier to β-sheet formation (9.6 ± 0.7 kJ/mol), whereas the barrier is near zero for turn formation. We emphasize that these are apparent barriers, because the Arrhenius analysis does not account for the temperature dependence of the viscosity of the solution. The apparent enthalpic barrier to β-sheet formation is likely due to the rearrangement of the hydrophobic interactions in the pre-collapsed state of the β-sheet of the hairpin. Although the activation energy is higher for the β-sheet than the turn, the observed rates are still faster, which suggests that the entropic barrier for turn formation is higher than for β-sheet formation. Taken together, these results support a model in which residual hydrophobic interactions at the terminus create a compact structure that has the correct registry for sheet formation, but still requires a search to find the stabilizing contacts and the correct alignment of the turn.

Computational studies of CLN025 reported multiple folding pathways for β-hairpin formation.[21] In the majority of pathways reported the proline in the turn reached the native state first and the sheet formed later. An alterative pathway where the residues in the sheet reach the native state prior to the turn is also observed. There is little difference in average folding time between the two observed pathways. Infrared T-jumps offer a way to validate theoretical models. Our experimental results on CLN025 support these computational studies by showing that there is no clear assignment of order of formation

but rather that turn and β-sheet formation occur competitively. This suggests that β-hairpin formation cannot be described as a simple turn zipper or hydrophobic collapse mechanism, but rather is a complex combination of these models.

2.7 Conclusion

The study of CLN025 reveals that a simple two-state model cannot be used to describe the formation of this β-hairpin and that the speed limit for folding of β-hairpins is considerably faster than previously hypothesized. Linear β-hairpins have been reported to fold in a range from 800 ns to over 10 μs. Folding of CLN025 occurs as fast as ~100 ns, a similar timescale to cyclic β-hairpins and to a-helical peptides. The ultrafast folding rate is a consequence of a negligible global barrier to folding and requires a recalibration of the speed limit for b-hairpin formation. Previous estimates of the speed limit for the formation of a small 10-residue interchain loop were ~10^6 s^{-1}.[24, 69] The relaxation lifetime observed for CLN025 breaks the speed limit by a factor of 10, and together with the heterogeneity in the kinetics indicates a smooth energy landscape with a minimal global barrier to folding. CLN025 folds as fast as α-helical peptides, demonstrating that β-hairpin formation may occur competitively with α-helical formation. Although a native-sequence β-hairpin that folds fast has not yet been discovered, this result at least provides proof of principle that β-hairpins can form quickly and therefore may act as nucleation sites for protein folding.

We propose a model to explain why CLN025 exceeds the previously proposed limit, based on residual interactions in the unfolded state between hydrophobic aromatic residues near the terminus. Such interactions lead to a pre-collapsed structure that only

requires local rearrangements to reach the folded state. Furthermore, different relaxation lifetimes are observed based on the position probed in the β-hairpin. A fast rate and small apparent activation energy are observed when the hydrophobic cluster at the terminus is probed. Conversely, when the turn is probed a slightly slower rate and negligible activation energy are observed. The heterogeneous kinetics support a folding model in which the rate-limiting step of β-hairpin formation is stabilization of the collapsed system over a minimal barrier to achieve the native turn alignment. The observation of a small apparent enthalpic barrier further justifies this idea as it implies that the rate is limited by the registry search and stabilization of turn contacts. In summary, we have observed ultrafast folding and distinct structural dynamics in the turn and the sheet of CLN025 that show the folding of a simple ten residue β-hairpin is heterogeneous and not a simple two state process. The highly optimized CLN025 sequence establishes a new speed limit for β-hairpin formation on a nearly barrierless free energy landscape.

Chapter 3: WW Domain Folding Complexity Revealed by Infrared Spectroscopy[70]

[Reproduced with permission from Davis, C. M., and Dyer, R. B. (2014) WW domain folding complexity revealed by infrared spectroscopy, *Biochemistry 53*, 5476-5484. http://doi.org/10.1021/bi500556h

3.1 Abstract

Although the intrinsic tryptophan fluorescence of proteins offers a convenient probe of protein folding, interpretation of the fluorescence spectrum is often difficult because it is sensitive to both global and local changes. Infrared spectroscopy offers a complementary measure of structural changes involved in protein folding, because it probes changes in the secondary structure of the protein backbone. Here we demonstrate the advantages of using multiple probes, infrared and fluorescence spectroscopy, to study the folding of the FBP28 WW domain. Laser induced temperature jumps coupled with fluorescence or infrared spectroscopy have been used to probe changes in the peptide backbone on the submillisecond timescale. The relaxation dynamics of the β-sheets and β-turn were measured independently by probing the corresponding IR bands assigned in the amide I region. Using these wavelength dependent measurements we observe three kinetics phases with the fastest process corresponding to the relaxation kinetics of the turns. In contrast, fluorescence measurements of the wild type WW domain and tryptophan mutants exhibit single exponential kinetics with a lifetime that corresponds to the slowest phase observed by infrared spectroscopy. Mutant sequences provide evidence for an

intermediate dry molten-globule state. The slowest step in the folding of this WW domain is the tight packing of the side chains in the transition from the dry molten globule intermediate to the native structure. This study demonstrates that using multiple complementary probes enhances the interpretation of protein folding dynamics.

3.2 Introduction

One of the most convenient probes of protein structure is the intrinsic fluorophore tryptophan. However, interpretation of tryptophan fluorescence is difficult because it can be sensitive to either local or global changes, or both. Furthermore, introduction of new tryptophan probes into a protein structure may perturb the structure and the folding dynamics and mechanism. A complementary approach is infrared spectroscopy of the amide-I mode of the peptide backbone. The advantage of using vibrational spectroscopy is that the spectral features are local; different regions of the amide-I' band report on different parts of a folded peptide or protein. Using infrared spectroscopy to independently probe the portion of the peptide assigned to different secondary structure elements it should be possible to determine the folding dynamics with greater resolution than by fluorescence spectroscopy.[71] β-proteins are naturally composed of two secondary structure elements, the turn and β-sheet. The turns are thought to act as nucleation sites for the folding of the larger β-protein.[72] By comparing the dynamics when the turn and β-sheet are probed, it should be possible to identify whether the turn and β-sheet form concurrently or with a specific ordering.

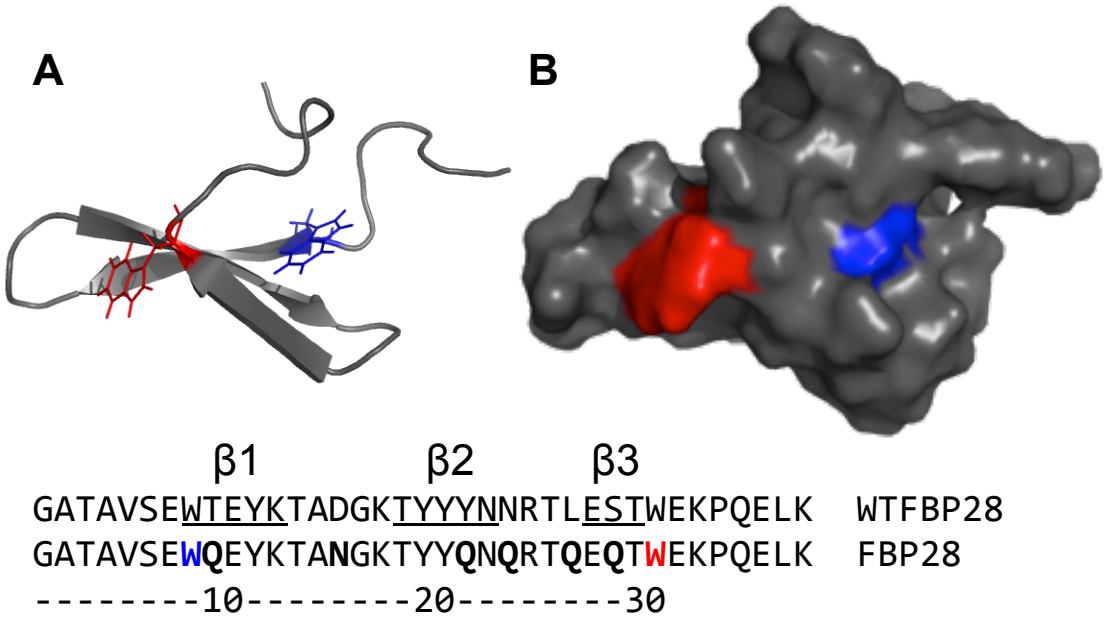


Figure 3-1. (A) Cartoon and (B) space filling model of FBP28 WW domain (PDB entry: 1E0L) showing the Trp8 (blue) and Trp30 (red) residues. Below: The mutated protein sequence, FBP28, compared to the wild type sequence, WTFBP28. The three β-sheets are labeled above and underlined in the WTFBP28 sequence. Mutations (bold), Trp8 (blue) and Trp30 (red) are highlighted in the FBP28 sequence. The figure was prepared in PyMOL (www.pymol.org).

Here we have compared time resolved measurements probed by fluorescence and infrared in one of the fastest folding β-protein families, the WW domain. The WW domain family consists of an antiparallel and highly twisted three-stranded β-sheet structure with a small hydrophobic core and two highly conserved tryptophan residues.[73-76] Because of the fast, submillisecond folding and native tryptophan fluorescence, WW domains have been the focus of extensive computational and experimental studies.[77-87]

Studies of the Formin Binding Protein 28 (FBP28) WW domain (Figure 3-1) have proposed different folding models involving either three-states[80, 81, 84, 88] or two-states.[79, 85, 86, 89] In many of the experimental studies laser induced temperature jumps probed by fluorescence were used to determine the folding kinetics.[79, 84-86] We employed two approaches to increase the structural resolution of the temperature jump measurements. In the first approach, infrared spectroscopy was used to independently probe the dynamics of the turn and β-sheet. In the second approach, tryptophan mutations were introduced to create a more localized fluorescent probe.

Mutation of the two conserved tryptophans in the FBP28 WW domain has been performed with varying success.[84, 86] The tryptophan in the second loop, Trp30, has been shown to be more amenable to mutation than the tryptophan in the first loop, Trp8, likely because Trp8 is buried in the hydrophobic core of the WW domain. Mutation of Trp8 to glycine or alanine completely destabilized the WW domain, whereas mutation to another aromatic residue, phenylalanine, still allowed folding into a WW domain.[86] Similarly, mutation of Trp30, located in the second loop, to phenylalanine had a slight stabilizing affect compared to mutation to alanine, which was slightly destabilizing.[84] Based on these observations, we concluded that mutation of the tryptophans to another aromatic group preserves aromatic packing necessary for the protein to fold. Therefore, we investigated fluorescent mutants in which the tryptophans are individually replaced with the aromatic residue tyrosine, W8Y and W30Y.

We compared the stability and folding kinetics of the wild type and mutated WW domains. The stability of the WW domains was determined by equilibrium circular dichorism and FTIR measurements. The dynamics of WW domain formation was

measured using temperature jump, time resolved infrared and fluorescence spectroscopy. Pulsed laser excitation was used to rapidly initiate a shift in the folding equilibrium. The relaxation of the WW domains was measured by both an infrared and a fluorescence probe. The dynamics of the β-sheets and turns were independently measured by probing the components of the IR amide I band assigned to each structure. We found that mutation of the Trp8 resulted in a destabilization of the peptide and a less rigid hydrophobic core. Because FBP28 W8Y does not fold with a tightly packed core, the slowest step in folding is eliminated in this case and therefore its folding rate is an order of magnitude faster than the other systems studied here. As expected, mutation of Trp30 had little affect on the protein's stability. Both mutants and the wildtype system exhibited multi-exponential behavior when probed by infrared spectroscopy, but only one phase when probed by fluorescence.

These experiments demonstrate the importance of using multiple probes to determine protein dynamics. Wavelength dependent infrared measurements reveal a fast 100 nanosecond phase, intermediate microsecond phase and slow hundreds of microseconds phase. These measurements provide direct evidence for a mechanism of WW domain formation initiated in the turns. The single phase observed by fluorescence coincides with the slowest phase observed by infrared spectroscopy. Tryptophan mutants enhance our understanding of the dynamics of the hydrophobic core. They suggest that there is an intermediate dry molten-globule state prior to the rearrangement of the tryptophan residues in the hydrophobic core of the WW domain. These results demonstrate the value of multiple spectroscopic probes in providing a more complete view of the mechanism of protein folding.

3.3 Experimental Section

Protein Synthesis and Purification. FBP28 (Figure 3-1) and tryptophan mutants were synthesized via standard 9-fluornylmethoxycarbonyl (Fmoc) based solid-phase chemistry on a Liberty1 microwave peptide-synthesizer (CEM, Matthews, NC). Fmoc-PAL-PS resin (Applied Biosystems, Foster City, CA) was used to form a peptide amide. The peptide was purified by reverse-phase chromatography (C18 column) using a water/acetonitrile gradient with 0.1% trifluoracetate (TFA) as the counter-ion. TFA interferes in the Amide-I IR measurements at 1672 cm^{-1}, so we remove it by anion exchange. The peptide was lyophilized and dissolved in a 2 mM HCl solution to allow exchange of the TFA counter-ion for chloride.[90] The identity of the peptide was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. The peptide was lyophilized and dissolved in D_2O to allow deuterium-hydrogen exchange of the amide protons. The peptide was lyophilized a second time and resuspended in D_2O buffer with 20 mM potassium phosphate buffer at pD* 7.0 (pD* refers to the uncorrected pH meter reading). Sample concentrations of 0.5-1.0 mM were prepared for both IR and fluorescence experiments.

CD Spectroscopy. CD spectra and CD melting curves were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Jasco temperature controller module (Jasco, Inc, Easton, MD). Peptides were dissolved at 50 μM in 20 mM potassium phosphate buffer at pH 7.0. All measurements were obtained using a 1-mm pathlength cell. Wavelength scans were recorded over the range of 260 nm to 190 nm with an average of 3 repeats. A bandwidth of 2 nm and a scan rate of 50 nm/min were used for spectral acquisition. Thermal unfolding experiments were performed by monitoring the

signal at 226 nm from 5 °C to 90 °C using a 0.1 °C interval and scan rate of 30 °C/hr. During the thermal unfolding experiment a full wavelength scan was obtained every 5 °C after a 60 second delay. The buffer and protein concentrations were the same as used in the wavelength scan experiment.

FTIR Spectroscopy. The equilibrium melting behavior was monitored on a Varian Excalibur 3100 FTIR spectrometer (Varian Inc., Palo Alto, USA) using a temperature controlled IR cell. The IR cell consists of two CaF_2 windows stacked and separated by a 100 μm Teflon spacer split into two compartments, a sample and a reference. The same cells are used for equilibrium FTIR and T-jump experiments. No aggregation was observed in the infrared at reported concentrations. All spectra shown at a specific temperature are constructed by subtracting the spectrum of reference buffer solution without protein from sample solution with protein. The temperature-dependent difference spectra were then generated by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures. The second derivative spectra were computed in IGOR PRO after smoothing the data with a sixth order binomial algorithm to remove any residual water vapor (WaveMetrics, Lake Oswego, OR).

Time Resolved Temperature Jump (T-jump) Relaxation Measurements. The IR T-jump apparatus has been described previously.[30] Pulsed laser excitation is used to rapidly perturb the folding equilibrium on a timescale faster than the molecular dynamics of interest. Time resolved infrared is then use to probe the reaction. A Q-switched DCR-4 Nd:YAG laser (Spectra Physics, Mountainview, CA) fundamental at 1064 nm is Raman shifted (one stokes shift in 200 psi H_2 gas) to produce a 10 ns pulse at 2 μm. The magnitude of the T-jump is calculated using the change in reference absorbance with

temperature. The T-jump reference is taken from D_2O buffer with 20 mM potassium phosphate buffer at pD* 7.0 at the same temperature and frequency as the sample. Absorbance changes at the reference frequency are due only to changes in D_2O absorbance, which is used as an internal thermometer.[30]

The change in signal induced by the T-jump is probed in real time by a continuous laser with a frequency in the amide I' band of the IR or at 285 nm to excite Trp fluorescence. The mid-IR probe beam is generated by a continuous wave quantum cascade laser (Daylight Solutions Inc., San Diego, CA) with a tunable output range of 1570-1730 cm^{-1}. The transient transmission of the probe beam through the sample is measured using a fast, 100 MHz, photovoltaic MCT IR detector/preamplifier (Kolmar Technologies, Newburyport, MA). Transient signals are digitized and signal averaged (1000 shots) using a Tektronics digitizer (7612D, Beaverton, OR).

The fluorescence excited at 285 nm and collected from 320-370 nm is sensitive to changes in the tryptophan residues. A Verdi V12 DPSS High-Power continuous wave laser (Coherent, Santa Clara, CA) is used to pump a Mira 900 Ti:Saph laser (Coherent, Santa Clara, CA) which produces a quasi-continuous beam at 855 nm. The beam is then passed through a second and third harmonic generator (Coherent, Santa Clara, CA) to produce the fluorescence probe beam at 285 nm. The back-emitted fluorescence light induced by the 285 nm pump laser is measured using a Hamamatsu R7518 photomultiplier tube (Hamamatsu Photonics K.K., Hamamatsu City, Japan), digitized and signal averaged (5000 shots) using a Tektronics digitizer (7612D, Veaverton, OR). A 400 μM tryptophan (Trp) solution in D_2O is used as a reference to measure the magnitude of the temperature jump and determine the temperature dependence of the tryptophan signal.

Instrument control and data collection are controlled using a LabVIEW computer program (National Instruments, Austin, TX).

Analysis of Kinetics Data. The peptide relaxation kinetics must be deconvolved from the observed kinetics. Accurate deconvolution is possible as the instrument response is determined from the reference measurement under the exact conditions of the sample measurements. In order to minimize detector artifacts, the reference is scaled prior to subtraction from the sample. The decay function is a triple exponential decay with the formula:

$$A = A_0 + \ldots + A_n \exp\left(\frac{-(x - x_o)}{\tau_n}\right) \qquad (3\text{-}1)$$

where A_0 is an offset, n is the number of exponentials to fit, A_n is a preexponential factor, τ_n is the relaxation lifetime of the sample and x_0 is the time offset. In order to best fit the data, the minimum number of exponentials with unique relaxation lifetimes was selected. The data are fit over the interval from 90 ns to an order of magnitude outside the slowest exponential. In the cases where a fast phase is not reported the fit starts at 400 ns. Fluorescence measurements are fit starting at 5 μs. The data analysis was performed in IGOR PRO.

Results and Discussion

3.4 Far-UV CD Spectroscopy. The CD spectra of the wild type FBP28 and, to a lesser extent, the FBP28 W30Y and FBP28 W8Y mutants exhibit a positive peak at 226 nm and a negative peak at 197 nm (Figure 3-2A). The typical β-sheet spectrum has a negative peak at ~218 nm and a positive peak at ~195 nm.[91] Folded WW domains instead have a

negative peak at ~202 nm and a positive peak at ~230 nm that provides a distinct CD signature.[92, 93] Small variations in the position and intensities of these peaks have been observed amongst the WW domain family.[93] The peak at ~202 nm resembles the random coil peak usually found at 200 nm. Disorder in the N- and C- termini of the folded structure of WW domains is thought to contribute to the negative peak in the CD spectrum.[92, 94-96] The peak at ~230 nm arises from the presence of ordered aromatic side chains.[79, 91, 92] Therefore, the CD spectra of FBP28 and the tryptophan mutants are consistent with a WW domain topology.

Mutation of the Trp8 or Trp30 results in a reduction in the intensity of the 226 nm peak. FBP28 W30Y exhibits a cooperative thermal denaturation detected by circular dichroism (Figure 3-2B) that closely resembles the wild type FBP28. This is further evidence that the W30Y mutant adopts a WW domain structure. Thus, the diminished 230 nm CD signal of the FBP28 W30Y mutant compared to the wild type FBP28 can be attributed to the absence of Trp30. A similar mutation to loop 2 of the human Pin1 WW domain was confirmed to be a WW domain based on the same observations, supporting this assignment.[97] FBP28 W8Y exhibits a weak feature at 230 nm with a broad, noncooperative dependence on temperature (Figure 3-2). WW-domains include a conserved Trp8-Tyr20-Pro33 hydrophobic core.[84] Apparently, mutation of Trp8 disrupts this core, affecting the packing of the entire domain. In this mutant, the observed CD band at 226 nm must be due to the Trp30. The expected contribution of Trp30 to the 226 nm CD signal for the correctly folded WW domain is obtained from the difference between the WT and W30Y mutant CD spectra shown in Fig. 3-2A. The weak feature at 226 nm observed for the W8Y mutant is not consistent with the expected contribution of

the Trp30 in the folded state. This suggests that the environment of the tryptophan in the outer loop is not rigid, supporting our observation that the fold is disrupted by the W8Y mutation.

Thermal denaturation was monitored by recording the absorbance change at 230 nm with temperature (Figure 3-2B). The WW domains exhibit the typical heat-induced unfolding behavior with a loss of intensity at 226 nm and a shift of the minimum at 197 nm to 200 nm, corresponding to a change in the secondary structure from WW domain to random coil (Figure 3-2A).[92] The melting curves (Figure 3-2B) were fit to an apparent two-state equilibrium model:

$$A_O = \frac{A_F}{1 + exp\left(-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_M}\right)\right)} + \frac{A_U}{1 + exp\left(\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_M}\right)\right)} \qquad (3-2)$$

where A_O is the observed absorbance, A_F and A_U are the absorbance contributions from the folded and unfolded populations, ΔH is the enthalpy change at the midpoint, R is the gas constant, and T_m is the transition midpoint.[98] This analysis assumes a ΔC_p of 0 as it is unlikely that a small peptide like a WW domain would have a large difference in heat capacity between the folded and unfolded states. The data are then normalized for comparison. The observed melting temperature of the wild type FBP28 is 51.1 ± 0.3 °C. This agrees with the previously reported melting temperature of FBP28, 50.9 °C, obtained by an infrared melt of the β-sheet band at 1636 cm^{-1}.[99] The melting temperature of the FBP28 W30Y mutant is 49.0 ± 0.1 °C. The W30Y mutation results in a decrease in the melting temperature of ~2 °C. The breadth of the transition of FBP28 and FBP28 W30Y are the same showing that while the mutation results in a slight destabilization, the cooperativity of the folding transition is unchanged (Figure 3-2B). The transition of

FBP28 W8Y was too broad to obtain a reliable fit. This is evidence that mutation of Trp8 results in destabilization of the fold such that the transition becomes less cooperative.

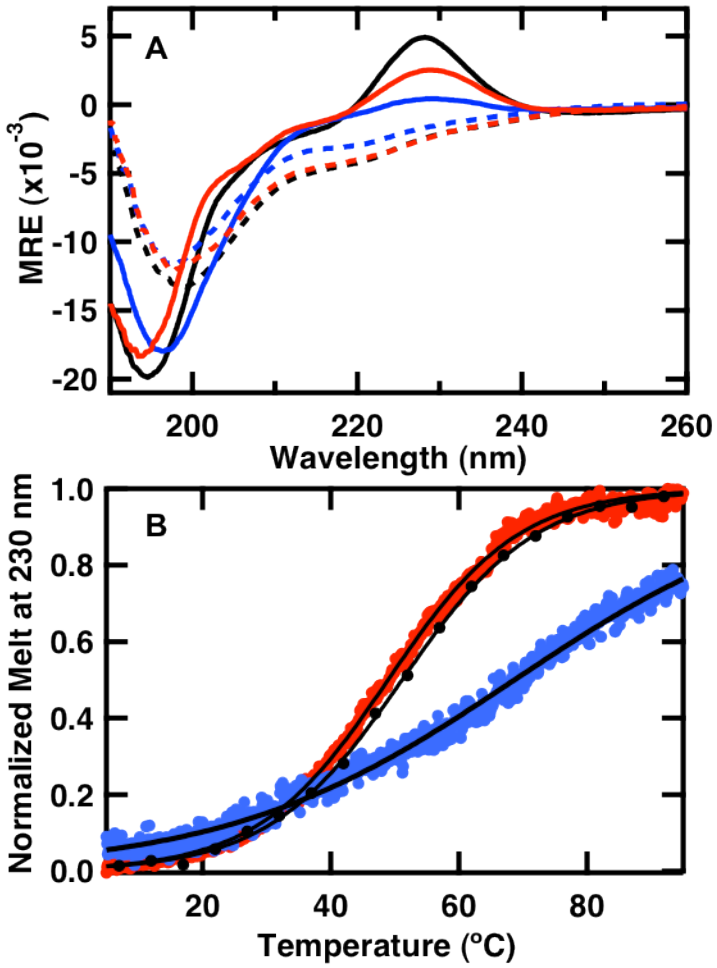


Figure 3-2. (A) Far-UV CD spectra of 50 μM solutions of FBP28 (black), FBP28 W8Y (blue), and FBP28 W30Y (red) WW domains in 20 mM potassium phosphate buffer (pD 7.0) acquired at 5°C (____) and 95 °C (----) during the course of a thermal denaturation in a 0.1 cm pathlength cell. (B) Thermal denaturation of the WW domains monitored by CD at 226 nm. The continuous line a fit to an apparent 2-state model (Equation 3-2) and then normalized.

3.5 FTIR Spectroscopy. The temperature-induced unfolding of FBP28 and the tryptophan mutants was studied over the range from 5 to 85 °C in 5 °C intervals using FTIR spectroscopy to monitor the amide I' region. An example of the temperature dependent absorption spectra of the amide I' spectral region (amide I region of peptides in D_2O) of FBP28 is shown in Figure 3-3A (see tryptophan mutants in Chapter 3 Appendix Figure 3A-1). The amide I' absorbance arises from C=O stretching vibrations of the polypeptide backbone carbonyls, and is an established indicator of secondary structure.[62-64] This relatively broad band contains contributions from the entire polypeptide backbone, which in the case of the WW domain includes β-sheet, β-turn, and random coil structure. The changes with temperature are highlighted by the difference spectra for each peptide (Figure 3-3B). The difference spectra are generated by subtracting the lowest temperature spectrum from each absorbance spectrum at higher temperature. Negative peaks correspond to specific structures or interactions present in the folded state, and positive peaks correspond to new interactions with solvent in the unfolded state. The individual peaks are more easily distinguished in the second derivative of the FTIR spectra at the lowest temperature (Figure 3-3C).

This analysis reveals three main components of the amide I' band at 5 °C, centered at 1613, 1638, and 1678 cm^{-1}. The intensity of these features decreases with increasing temperature, meaning that they are all associated with the folded state. These peaks have previously been observed in other WW domains.[88, 100] A peak at ~1611 cm^{-1} in β-hairpins has been assigned to an amide C=O group in the turn usually involved in multiple hydrogen bonds with side chain or backbone donors.[50, 66] There is one such group in the first turn and none in the second turn of the wildtype FBP28 WW domain.[74, 101] We assign

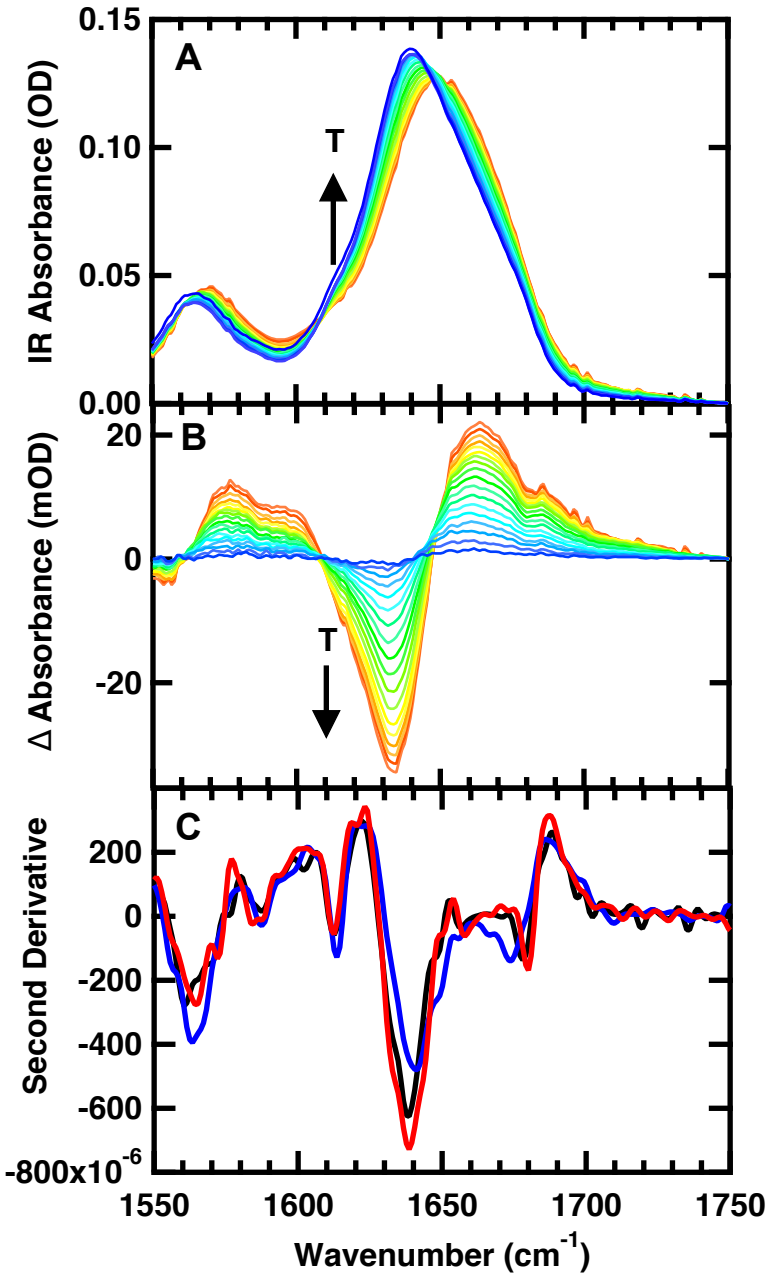


Figure 3-3. Temperature dependent FTIR spectra of 1 mM FBP28 in 20 mM potassium phosphate buffer (pH 7). (A) Absorbance spectra in the Amide I' region; the temperature of the individual traces varies from 5 to 85 °C in 5 °C intervals. (B) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher temperatures. (C)

Second derivative of the FTIR absorbance spectrum of FBP28 (black), FBP28 W8Y (blue), and FBP28 W30Y (red) WW domains at 5 °C.

the peak at 1613 cm^{-1} to the backbone amide I' mode of threonine 13, which is strongly hydrogen-bonded to the backbone N-Hs of glycine 16 and lysine 17 in the first turn of the WW domains. This peak is of equal intensity in all three peptides (Figure 3-3C) as all three peptides contain the same first turn. IR bands at 1634 and 1681 cm^{-1} are well-established components of antiparallel β-sheets.[102] The amide I band is split into a peak at 1681 cm^{-1} and 1634 cm^{-1} due to out-of-phase and in-phase C=O stretching of neighboring amides. The efficiency of the interstrand coupling depends on the orientation of the dipole moments of the individual carbonyl oscillators. Misaligned carbonyl oscillators do not couple as efficiently leading to a decrease in splitting. The splitting is sensitive to the number of strands in a β-sheet; as the number decreases, the amide I splitting and intensity also decrease.[102] The FBP28 and FBP28 W30Y amide I' bands at 1638 and 1678 cm^{-1} are consistent with these characteristic β-sheet markers. The β-sheet peaks of the FBP28 W8Y mutant have decreased intensity and splitting, such that they are shifted to 1641 and 1673 cm^{-1} (Figure 3-3C). One interpretation of the FBP28 W8Y data is that the decrease in splitting and intensity of the in and out-of-phase C=O bands is due to the formation of fewer strands (e.g. a 2-stranded β-sheet) compared to the wild type system. Infrared and CD measurements do not support this model. The peak at 1613 cm^{-1} is due to the turn of the first loop, indicating that the turn is formed and the strong cross-strand hydrogen-bonds are in place. CD measurements of FBP28 W8Y at low temperature show that the Trp30 resides in a rigid environment, so there is also structure to the third β-sheet

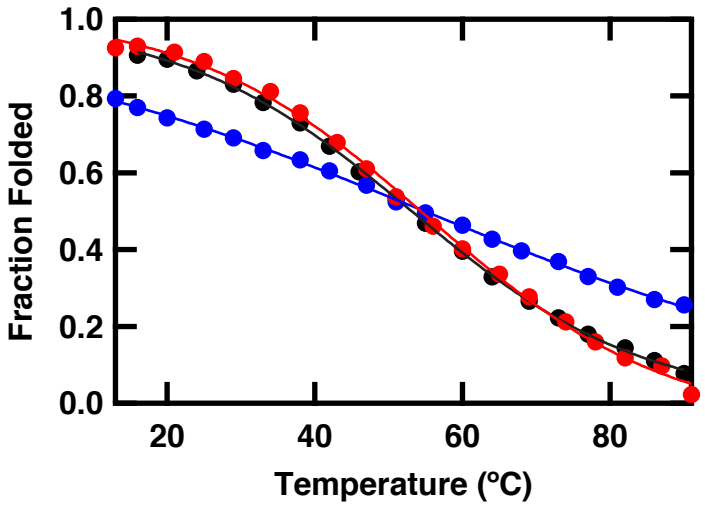


Figure 3-4. FTIR melt curves for FBP28 (black), FBP28 W8Y (blue) and FBP28 W30Y (red) WW Domains obtained by plotting the change in IR difference spectra at 1634 cm^{-1} versus temperature. The data are fit to an apparent 2-state model (Equation 3-2) and then normalized.

and the second loop. These results demonstrate that there is structure in both the first and second loop, so the decrease in splitting is not due to loss of a β-sheet. Instead, we postulate that disruption of the hydrophobic core caused by the Trp8 mutation misaligns the carbonyl oscillators in the three sheets causing the decrease in splitting and intensity.

The normalized melting curves for the WW domains derived from the temperature dependent IR absorbance at 1634 cm^{-1} are shown in Figure 3-4. The data were normalized after being fit to an apparent two-state equilibrium model using equation 3-2. The melting curve probed by infrared, Figure 3-4, is more broad than the melting curve probed by CD, Figure 3-2B, likely because infrared probes the secondary structure of the entire peptide, whereas, CD probes the chirality of the environment of the tryptophan

residues. The melting temperatures derived from the amide I' absorbance of FBP28, 52 ± 1°C, and FBP28 W30Y, 53 ± 1°C, are the same within error. The melting curve derived from other wavelengths agreed with the melting temperature within the error of the fit. That the melt and breadth of transitions of the FBP28 and FBP28 W30Y mutant agree supports the conclusion from the circular dichroism measurements that the Trp30 mutation does not affect the WW domain fold. The transition of FBP28 W8Y was too broad to obtain a reliable fit. This supports the observation from the FTIR and circular dichroism analysis that the W8Y mutation destabilizes the fold.

3.6 Temperature-Jump Relaxation Kinetics. The probe dependent relaxation kinetics of the folding/unfolding transition following a laser induced temperature jump were monitored using time-resolved infrared spectroscopy. Structure specific measurements were made using the amide I' frequency for the amide bond vibration of the turn (1619 cm^{-1}) and residues involved in coupling between the strands of the β-sheet (1634 cm^{-1}). Jumps were performed slightly off peak center to maximize the transient absorbance signal. The complete relaxation kinetics for each peptide is reported in the appendix (Appendix Table 3A-1, 3A-2). Time-resolved measurements examined the dependence of the relaxation rates on the final temperature following a temperature jump. The magnitude of the temperature jump was kept constant while varying the final temperature for a range of final temperatures below the melting transition. The relaxation kinetics is best fit by a triple exponential (Equation 3-1, Appendix Figure 3A-2). Figure 3-5 displays the relaxation kinetics of the peptides following a jump from 15 to 30 °C. The data were normalized for comparison. The fits of the data in Figure 3-5 are reported in Table 3-1.

There is good agreement between the dynamics of the FBP28 and FBP28 W30Y mutant (Appendix Table 3A-2). Both exhibit a fast 100's of ns phase (τ_1), a μs phase (τ_2), and a slower ~100 μs phase (τ_3). FBP28 W8Y also exhibits three phases, with the two slower phases an order of magnitude faster than the other peptides. T-jump IR relaxation transients were also collected at 1661 cm^{-1} at all temperatures probed, and they exhibit three kinetic phases with the same lifetimes as those observed at 1619 and 1634 cm^{-1}. The 1661 cm^{-1} probe frequency is sensitive to disorder in the peptide, whereas the probes at 1619 and 1634 cm^{-1} report directly on specific secondary structures. Because of this, the data at 1661 cm^{-1} were useful as a control, but do not provide any new information about the folding of specific secondary structures within the WW domain.

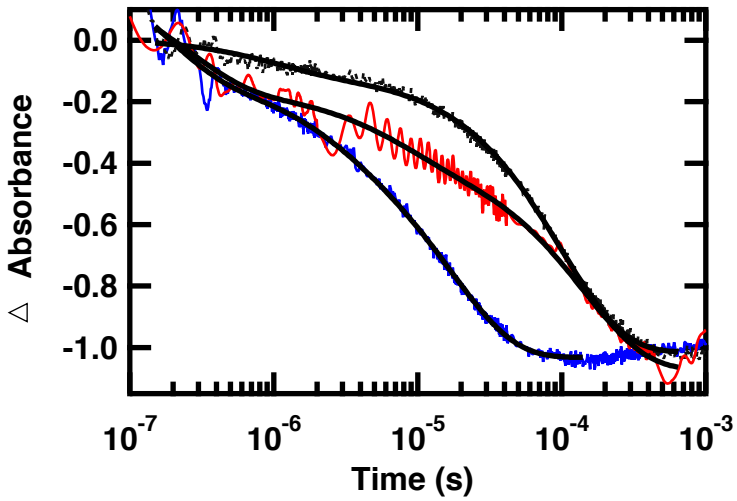


Figure 3-5. Representative IR T-jump relaxation kinetics of FBP28 (black), FBP28 W8Y (blue) and FBP28 W30Y (red) monitored in the amide I' spectral region at 1634 cm^{-1} following a T-jump from 15 to 30 °C. A triple exponential fit using equation (3-1) is overlaid on each kinetic trace. The data are normalized at the minimum for comparison.

Table 3-1: Relaxation Kinetics Following a Temperature Jump to ~30 °C

	Infrared (1634 cm^{-1})						**Fluorescence**
	A_1 (%)	τ_1 (ns)	A_2 (%)	τ_2 (μs)	A_3 (%)	τ_3 (μs)	τ (μs)
FBP28	11 ± 4	230 ± 15	6 ± 4	9 ± 1	83 ± 1	105 ± 1	88 ± 5
FBP28 W8Y	20 ± 4	140 ± 12	15 ± 1	2.1 ± 0.2	65 ± 1	17.3 ± 0.3	22 ± 6
FBP28 W30Y	18 ± 2	320 ± 23	19 ± 2	7.4 ± 0.5	63 ± 1	143 ± 3	77 ± 4

Wavelength dependent measurements reveal differences in the dynamics of the turn of loop 1 (1619 cm^{-1}) and the β-sheet of the WW domain (1634 cm^{-1}) (Figure 3-6). The relaxation lifetimes observed at 1619 cm^{-1} and 1634 cm^{-1} are similar (Table 3A-1, 3A-2). Because there is overlap between the peaks at 1619 and 1634 cm^{-1}, probes at either location are sensitive to dynamics associated with folding at both locations, resulting in multi-exponential kinetics regardless of the probe frequency. The relative amplitude of each kinetics phase depends on the probe frequency, however, which allows us to assign the structural feature that contributes most strongly to each phase. The relative amplitude of the fast phase is greatest in the transient measured at 1619 cm^{-1}, which probes the turn of loop 1. In contrast, the slower phases dominate the kinetics measured at 1634 cm^{-1}, which probes interstrand coupling across the sheet of the WW domain (Figure 3-6). Based on these observations we assign the fast phase to turn formation in loop 1, whereas the slow phases are assigned to sheet formation. This conclusion supports the prediction from previous studies that FBP28 forms through an intermediate state where the first

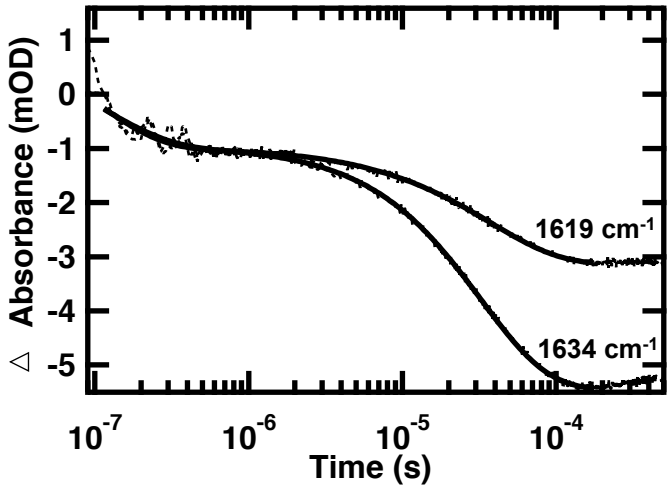


Figure 3-6. Representative IR T-jump relaxation kinetics of FBP28 monitored in the amide I' spectral region at 1619 and 1634 cm^{-1} following a T-jump from 15 to 30 °C. A triple exponential fit is overlaid on each kinetic trace (black solid line).

hairpin is highly structured.[78, 86, 103]

In previous work, measurements of FBP28 WW domain with the fast folding hairpin chignolin incorporated into the loops resulted in similar triple exponential kinetics, where the intermediate phase is interpreted as arising from the development of the cross-strand interactions of loop 1 to form the first half of the WW domain.[88] Chignolin is a 10 residue hairpin whose sequence was optimized for stability based on the 16 residue β-hairpin G-peptide.[55] Replacing a native WW domain turn with the chignolin sequence created a difference in stability of the two loops, which resulted in characteristic IR peaks that could be assigned to each of the loops. This made it possible to probe formation of the first and second loop separately. Because the folding of the two hairpins in the FBP28 WW domain and mutants is more cooperative, it is not possible to independently probe

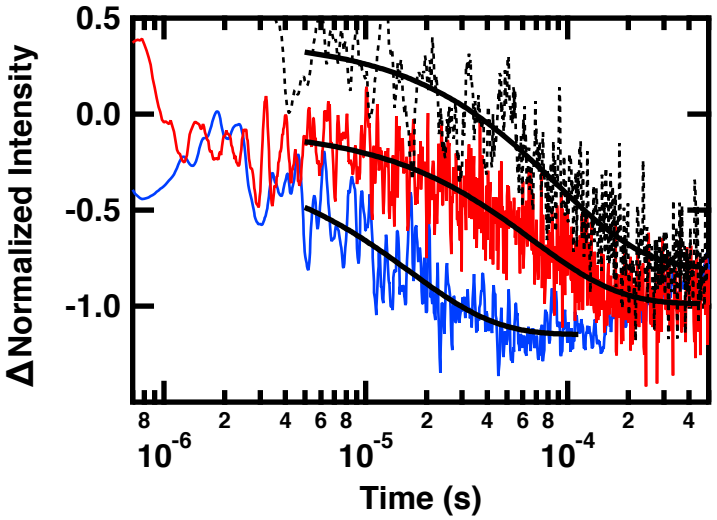


Figure 3-7. Representative fluorescence T-jump relaxation kinetics of FBP28 (black), FBP28 W30Y (red) and FBP28 W8Y (blue) excited at 285 nm and monitored at 350 nm following a jump from 20 to 28 °C. Single exponential fits are overlaid on the kinetics traces. Data are offset for clarity.

the dynamics of the first and second loop; however, the intermediate phase likely arises from formation of the cross-strand interactions of the first loop, since turn 1 is already formed and by analogy to the chignolin containing system. These results support a model in which the first hairpin is formed in the intermediate state.

A second set of kinetics experiments with a series of similar magnitude T-jumps was performed using time-resolved fluorescence as a probe. Figure 3-7 shows the relaxation kinetics for a T-jump from 20 to 28 °C probed by tryptophan (Trp) fluorescence. The fits of the data in Figure 3-7 are reported in Table 3-1. The complete relaxation kinetics for each peptide and a figure demonstrating the full resolution of the fluorescence T-jump are reported in the appendix (Figure 3A-3, Table 3A-3). In WT and each of the mutants we

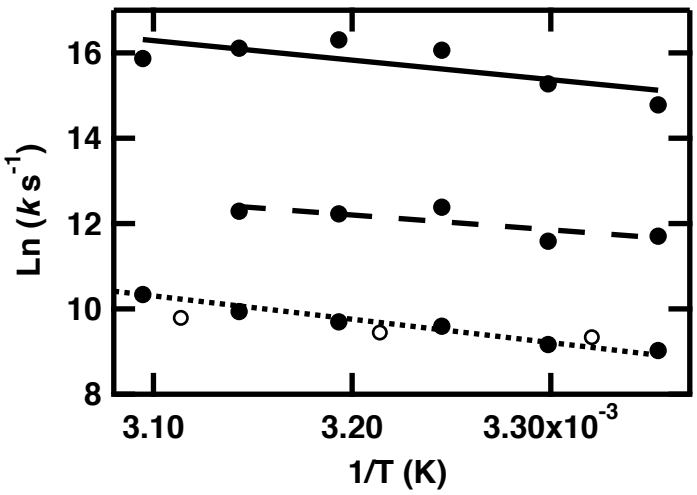


Figure 3-8. Arrhenius plot of the relaxation kinetics of FBP28 in the folding branch (below T_m). The values of T used for the (1/T) axis are the final temperatures reached during the jump. *k* is the value obtained from a fit (see text) of the T-jump transient probed by infrared at 1634 cm^{-1} (●) or fluorescence (○) spectroscopy. Lines are a result of fitting τ_1 (solid line), τ_2 (dashed line), and τ_3 (dotted line).

observe a single phase in the fluorescence relaxation kinetics, corresponding to the slowest IR probed relaxation lifetime; we have already assigned the latter to folding of the second strand of the β-sheet (Figure 3-8). Since the fluorescence kinetics coincide with the slowest IR phase, both report on the final step in the folding process. When the fluorescence probe is in the first loop (the W30Y mutant has a single Trp reporter, W8, in the first loop), the observed fluorescence change corresponds to the ~100 μs step. When the Trp is on the second loop (W8Y), the fluorescence transient corresponds to the ~10 μs step, but there is no slower phase in this case.

The Trp fluorescence does not report on the peptide backbone conformation (at least not directly), but rather the packing of the Trp side chains in the hydrophobic core. The Trp side chains are at least partially buried in the hydrophobic core based on the FBP28 structure (Figure 3-1), so the Trp fluorescence maximum reports on the formation of this core. The fluorescence spectra of FBP28 and FBP28 W30Y have similar maxima near 345 nm, typical of Trp buried in a hydrophobic environment. In contrast, the FBP28 W8Y has a decreased intensity and red shifted maximum at 360 nm (Figure 3A-4), consistent with a more solvent exposed Trp. The fluorescence spectrum of each of the Trp mutants is sensitive to the changes in the environment of the single remaining Trp, indicating that W8 is solvent protected as expected from the structure (Figure 3-1) whereas W30 is more solvent exposed, causing the red shift of the W8Y fluorescence spectrum. The solvent exposure of W30 is likely increased in the W8Y mutant due to disruption of the hydrophobic core, consistent with evidence from the IR and CD spectra. Regardless, in all cases the fluorescence kinetics can be assigned to formation of the hydrophobic core and at least partial protection of the indole ring (or rings in the case of FBP28) from solvent. The correspondence of the fluorescence kinetics with the slowest IR phase indicates that burial of the Trp occurs concomitant with the folding to the native backbone topology. Therefore, the formation of the hydrophobic core depends on acquiring the native topology, meaning there is not an initial nonspecific hydrophobic collapse followed by a rearrangement to correctly align the three strands of the β-sheet structure. Importantly, this process happens faster for the W8Y mutant, which does not form a tightly packed hydrophobic core.

An intermediate dry molten-globule state in which solvent has been expelled from the protein core, but having only loosely packed side chains, has been observed in several proteins.[104-109] A dry molten-globule state has been observed in a similarly sized protein to the FBP28 WW domain, the 35 residue villin headpiece subdomain (HP35).[110] The FBP28 W8Y mutant folds to a final state that has all of the characteristics of a dry molten globule. Its IR spectrum is consistent with a native-like WW domain topology, having the turns and stabilizing cross-strand interactions that make up the β-sheet. Its fluorescence spectrum indicates a solvent protected Trp30. But the critical observation is its CD spectrum, which does not have the characteristic Trp peak at 230 nm, indicating that although the Trp is buried in the core, it is disordered. We conclude that the hydrophobic core is disrupted by the Trp8 mutation and therefore does not pack tightly. The IR T-jump data provide further support for this model because FBP28 W8Y exhibits simplified folding kinetics, missing the slowest ~100 μs phase observed for wild type, and the fluorescence only exhibits the ~10 μs phase. These results are consistent with a model in which the W8Y mutant folds to a dry-molten globule state, but the side chains are unable to tightly pack into the final folded structure. Hydrophobic collapse to the dry molten globule state occurs on the same timescale of formation of the second loop (~10 μs). Because the FBP28 W8Y mutant does not have to search for the final close-packed side chain conformations, the folding is an order of magnitude faster than WT FBP28.

Comparison of the folding kinetics of the wild type sequence with the W8Y mutant implies that the slow (100 μs) step in the folding of the former is tight packing of the side chains to form the native structure. Similarly, the W30Y mutant clearly folds to a native-like structure (having the Trp 230 nm CD band and IR and fluorescence spectra nearly

identical to wild type) with the same kinetics as the wild type sequence. Folding of the W8Y mutant occurs on the same timescale as the intermediate phase observed in the folding of the parent FBP28 and the W30Y mutant. Therefore we assign the ~10 μs phase to formation of a dry-molten globule state and the ~100 μs phase to the conversion of the dry molten globule to the tightly packed native structure as the rate limiting step. Two different folding models involving the dry molten globule have been proposed within various systems. In the 'dry molten globule hypothesis' the rate limiting step is packing of the side-chains into their native conformation from the dry molten globule.[106, 107, 109] In the other model, expulsion of water from the hydrophobic core to form the dry molten globule is the rate-limiting step.[104-106, 108, 110] Our observations support the first mechanism for FBP28, for which the tight packing of the side chains is the rate-limiting process.

Previous fluorescence temperature-jump studies of FBP28 revealed biexponential kinetics below 65 °C with relaxation lifetimes of ~30 μs and >900 μs.[84] In those studies a single point mutation, W30F, was shown to change the kinetics from a biexponential to single exponential. The interpretation of this was inconclusive, as the change could either be a reporter effect or indicate a change in folding mechanism. We did not observe the slow >900 μs phase in our fluorescence temperature-jump measurements. A major difference in the present measurements is that we modified to FBP28 sequence to include point mutations designed to limit aggregation. The slow phase may have been lost due to sensitivity of the observed kinetics to aggregation. The slow phase was also not observed in extensive studies performed in the Fersht group.[79, 85, 86] They concluded that the slow phase may be an early step in off pathway aggregation, but not part of the folding mechanism. Since the FBP28 mutant that we studied was designed to eliminate

aggregation, we would not expect to observe this phase.[99] The fluorescence measurements that we performed were consistent with the fast ~30 μs phase that was previously observed. Infrared measurements offer a level of detail not available from the fluorescence experiments. In particular, frequency dependent IR measurements reveal fast nanosecond and microsecond phases not captured by fluorescence measurements. Because the infrared measurements are sensitive to changes in the peptide backbone, they are able to capture subtle changes and do not rely on a side chain reporter. This highlights the local sensitivity of infrared spectroscopy over more conventional fluorescence probes.

3.7 Conclusion

Using infrared spectroscopy we are able to observe subtle differences in the timescale of formation of the turns and β-sheets of the WW domain. We propose a model to describe WW domain folding initiated in the turns, similar to that proposed for other WW domains.[78, 86, 88, 103] Formation of the first turn initiates the formation of a two stranded sheet from the first loop. Subsequently, the second turn and third strand fold to form the full WW domain structure. Fluorescence T-jump was unable to observe the level of detail available from infrared spectroscopy. Whereas infrared spectroscopy is sensitive to changes in secondary structure of the protein, fluorescence spectroscopy is sensitive to the formation of the hydrophobic core and thus the measurements are complementary. Tryptophan mutants provide evidence for an intermediate dry molten-globule state. We observe packing of the tryptophan residues on the same timescale of formation of the second loop. Thus, by utilizing multiple probes we gain a more detailed picture of the folding landscape.

3.8 Appendix

1. Temperature dependent FTIR spectra of FBP28 W8Y and FBP28 W30Y were recorded to confirm that the peptide was folded. Temperature dependent FTIR of FBP28 W8Y and FBP28 W30Y are reported in Figure 3A-1.

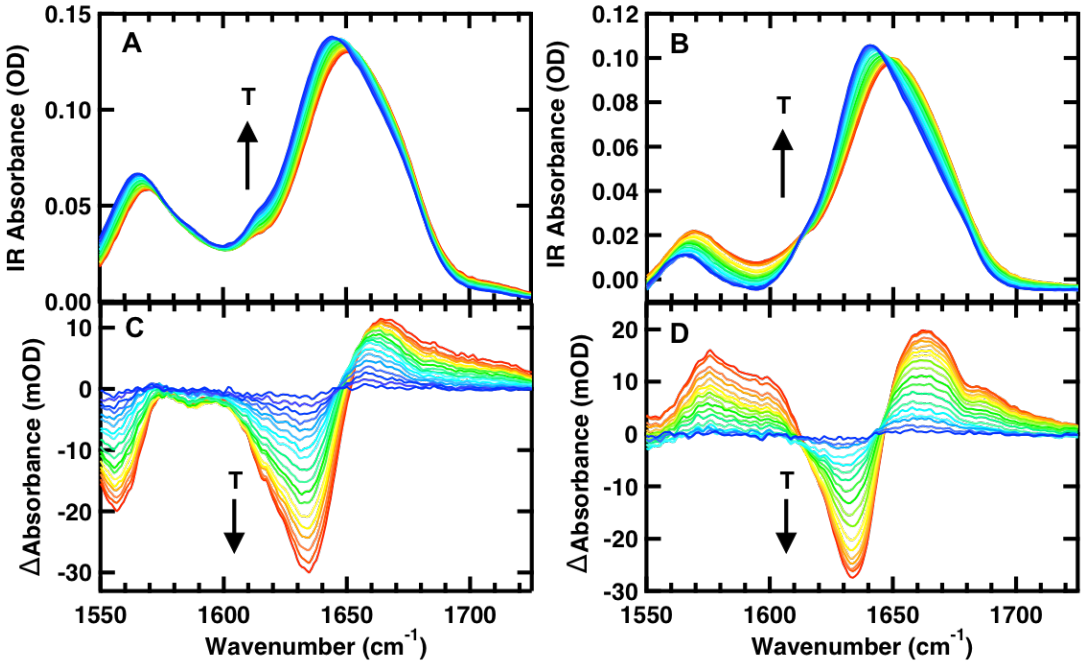


Figure 3A-1. Temperature dependent FTIR spectra of 1 mM FBP28 W8Y (A, C) and FBP28 W30Y (B, D) in 20 mM potassium phosphate buffer (pH 7). (A, B) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 5 to 85 °C in 5 °C intervals. (C, D) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher temperatures.

2. Temperature dependent infrared and fluorescence T-jump measurements were collected to determine the dynamics of the systems. Wavelength dependent IR T-jump

measurements were collected at 1619 (Table 3A-1) and 1634 cm^{-1} (Table 3A-2). 15 °C jumps were performed with an initial temperature from 10 to 35 °C. Data reported includes the observed magnitude and relaxation lifetime of each event. A comparison of the residual from a single and multi-exponential fit reveals that a single exponential cannot fit the data (Figure 3A-2). Fluorescence T-jump measurements are reported in Table 3A-3. ~10 °C jumps were performed with an initial temperature from 10 to 40 °C. Data reported include the relaxation lifetime of each event. The full time response of the Fluorescence T-jump instrument is demonstrated for FBP28 in Figure 3A-3. Each trace is the average of 5000 shots. The data are fit to a single exponential with no evidence of the two faster phases observed by IR T-jump.

Table 3A-1: Relaxation Kinetics Probed in the Turn of Loop 1 (1619 cm^{-1}) of FBP28, FBP28 W8Y, FBP28 W30Y

FBP28						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
10-25	0.28 ± 0.01	0.46 ± 0.01	1.84 ± 0.01	260 ± 26	5.2 ± 0.3	133 ± 2
15-30	0.49 ± 0.02	0.27 ± 0.01	1.68 ± 0.01	150 ± 11	4.1 ± 0.4	75 ± 1
20-35	0.61 ± 0.03	0.22 ± 0.01	1.73 ± 0.01	106 ± 8	3.7 ± 0.6	66 ± 1
25-40	0.48 ± 0.01	0.18 ± 0.02	1.89 ± 0.02	116 ± 7	5 ± 1	59 ± 1
30-45	0.55 ± 0.02	0.26 ± 0.01	2.13 ± 0.01	71 ±5	2.9 ± 0.3	50 ± 1
35-50	0.79 ± 0.04	0.13 ±0.03	1.99 ±0.03	108 ± 8	4.9 ± 1.8	38 ± 1
FBP28 W8Y						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
10-25	0.74 ± 0.03	0.85 ± 0.03	1.02 ± 0.01	130 ± 10	1.0 ± 0.5	19.2 ± 0.4
15-30	0.60 ± 0.03	0.65 ± 0.03	0.85 ± 0.01	100 ± 11	0.9 ± 0.1	16.9 ± 0.6
20-35	0.70 ± 0.03	0.39 ± 0.03	0.59 ± 0.02	130 ± 11	1.1 ± 0.2	11.4 ± 0.7
25-40	1.06 ± 0.04	0.41 ± 0.03	0.25 ± 0.03	111 ± 8	1.3 ± 0.2	14 ± 3
30-45	1.05 ± 0.03		0.31 ± 0.01	91 ± 4		5.0 ± 0.4
FBP28 W30Y						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
20-35		0.47 ± 0.05	0.62 ± 0.04		13 ± 2	150 ± 36
25-40		0.62 ± 0.03	0.89 ± 0.02		8.6 ± 0.9	130 ± 12
30-45	0.55 ± 0.04	0.55 ± 0.06	0.84 ± 0.06	180 ± 24	14 ± 2	95 ± 9
35-50	0.64 ± 0.07	0.31 ± 0.03	1.12 ± 0.03	90 ± 16	2.3 ± 0.6	43 ± 3

Table 3A-2: Relaxation Kinetics Probed in the Sheet of WW Domain (1634 cm^{-1}) of FBP28, FBP28 W8Y, FBP28 W30Y

FBP28						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
10-25	0.30 ± 0.02	0.37 ± 0.01	3.37 ± 0.01	380 ± 46	8.3 ± 0.8	121 ± 1
15-30	0.46 ± 0.02	0.25 ± 0.01	3.45 ± 0.01	230 ± 15	9.3 ± 1.1	105 ± 1
20-35	0.64 ± 0.03	0.30 ± 0.01	4.02 ± 0.01	106 ± 8	4.2 ± 0.5	68 ± 1
25-40	0.49 ± 0.03	0.33 ± 0.01	3.88 ± 0.01	83 ± 7	4.9 ± 0.5	62 ± 1
30-45	0.59 ± 0.03	0.30 ± 0.02	4.52 ± 0.02	101 ± 8	4.6 ± 0.7	48 ± 1
35-50	0.56 ± 0.02		4.48 ± 0.01	129 ± 8		33 ± 1
FBP28 W8Y						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
10-25	0.70 ±0.02	0.51 ± 0.02	2.24 ± 0.01	160 ± 11	1.4 ± 0.1	19.8 ± 0.2
15-30	0.68 ± 0.03	0.50 ± 0.03	2.21 ± 0.03	140 ± 12	2.1 ± 0.2	17.3 ± 0.3
20-35	0.56 ± 0.07	0.18 ± 0.07	1.75 ± 0.05	180 ± 39	1.1 ± 0.7	10.4 ± 0.3
25-40	0.48 ± 0.06		1.27 ± 0.01	210 ± 29		6.4 ± 0.2
30-45	0.71 ± 0.04		0.48 ± 0.01	140 ± 13		5.0 ± 0.4
35-50	0.57 ± 0.04		0.86 ± 0.01	160 ± 36		3.3 ± 0.3
FBP28 W30Y						
T-jump (°C)	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	0.42 ±0.01	0.44 ± 0.01	1.45 ± 0.01	320 ±23	7.4 ± 0.5	143 ± 3
20-35	1.98 ± 0.07	1.19 ± 0.03	2.02 ± 0.02	131 ± 9	3.2 ± 0.2	138 ± 6
25-40	2.0 ± 0.1	1.40 ± 0.08	2.97 ± 0.03	120 ± 16	1.9 ± 0.2	107 ± 4
30-45	0.78 ± 0.06	1.20 ± 0.04	3.91 ± 0.02	120 ± 20	1.8 ± 0.1	62 ± 1
35-50	0.54 ± 0.07	0.56 ± 0.06	5.41 ± 0.03	180 ± 46	2.2 ± 0.1	51 ± 1

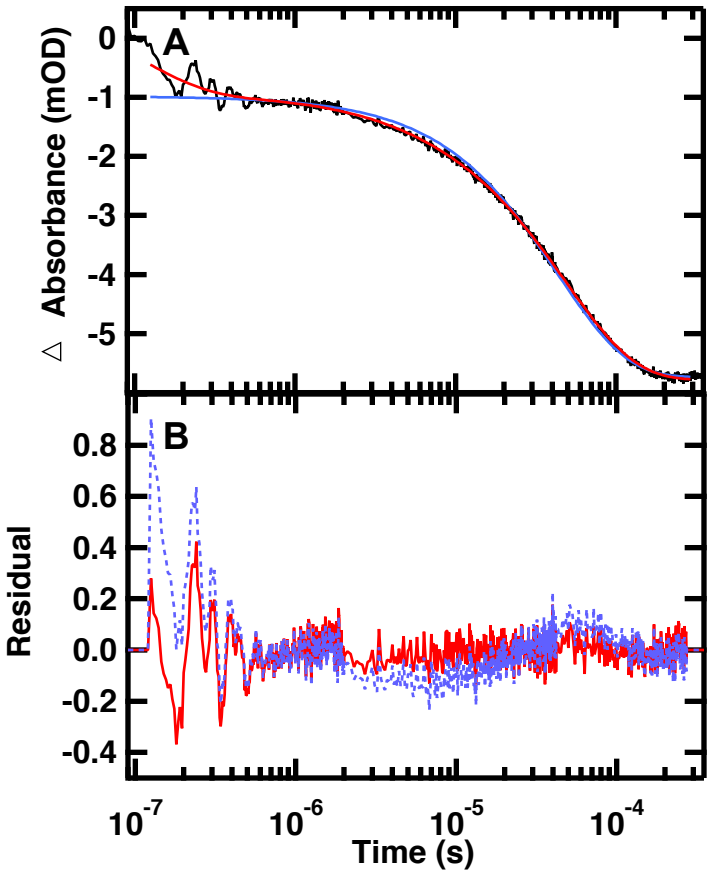


Figure 3A-2. (A) IR T-jump relaxation kinetics of FBP28 monitored in the amide I' spectral region at 1634 cm^{-1} following a T-jump from 30 to 45 °C. A single (blue) and triple (red) exponential fit is overlaid on the kinetic trace. (B) Comparison of residual from single exponential fit (blue) and triple exponential fit (red).

Table 3A-3: Relaxation Kinetics Probed by Tryptophan Fluorescence of FBP28, FBP28 W8Y and FBP28 W30Y

FBP28	
T-jump ΔT(Ti-Tf)/°C	τ (μs)
20-28	88 ± 5
30-38	79 ± 2
40-48	56 ± 1
FBP28 W8Y	
T-jump ΔT(Ti-Tf)/°C	τ (μs)
15-23	45 ± 3
20-28	22 ± 6
25-33	15 ± 7
30-38	10 ± 1
35-43	4 ± 1
FBP28 W30Y	
T-jump ΔT(Ti-Tf)/°C	τ (μs)
10-18	82 ± 2
20-28	77 ± 4
30-38	53 ± 3
40-48	25 ± 1

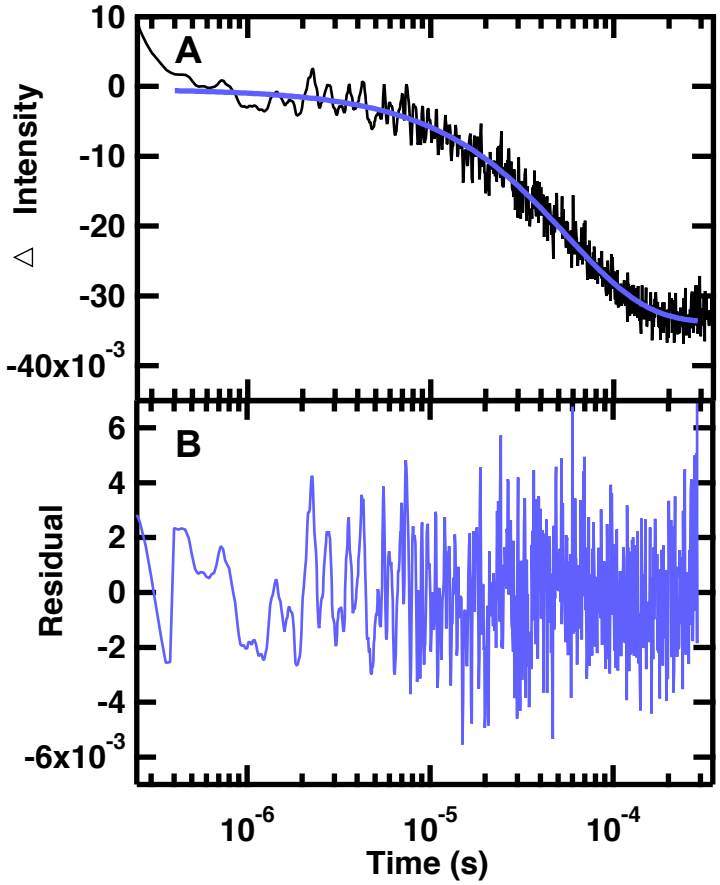


Figure 3A-3. (A) Fluorescence T-jump relaxation kinetics of FBP28 excited at 285 nm and monitored at 350 nm using a 200 Ω resistor following a jump from 40 to 48 °C. A single exponential fit is overlaid on the kinetic trace. (B) Residual from single exponential fit.

3. Equilibrium Fluorescence of FBP28, FBP28 W8Y and FBP28 W30Y were collected to compare formation of the hydrophobic core. Fluorescence spectra were recorded on a HORIBA Jobin Yvon FluoroMax-3 Spectrofluorometer (HORIBA Scientific, Edison, NJ). Peptides were dissolved at 10 μM in 20 mM potassium phosphate buffer at pH 7.0. All measurements were obtained using a 1.0 cm pathlength cuvette. The peptide was excited at 280 nm and monitored from 300 nm to 450 nm with an average of 3 repeats. Equilibrium fluorescence measurements are reported in Figure 3A-4.

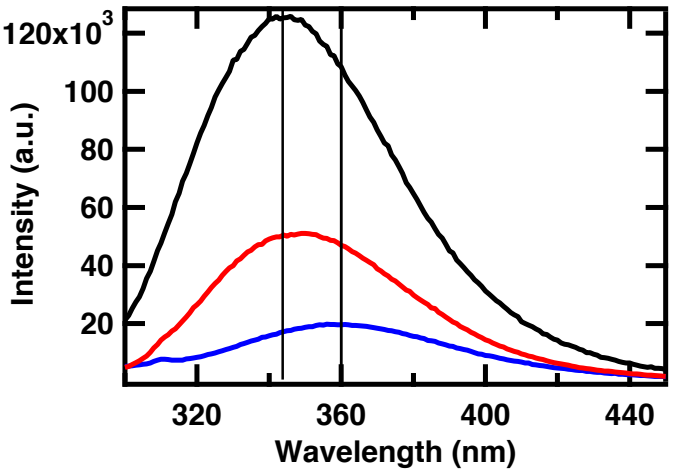


Figure 3A-4. Comparison of fluorescence spectra of FBP28 (black trace), FBP28 W8Y (blue trace) and FBP28 W30Y (red trace) excited at 280 nm. Vertical lines indicate the peak maximum of FBP28 and FBP28 W8Y.

Chapter 4: Dynamics of an Ultrafast Folding Subdomain in the Context of a Larger Protein Fold[88]

4.1 Abstract

Small fast folding subdomains with low contact order have been postulated to facilitate the folding of larger proteins. We have tested this idea by determining how the fastest folding linear β-hairpin, CLN025, which folds on the nanosecond timescale, folds within the context of a two-hairpin WW domain system, which folds on the microsecond timescale. The folding of the wild type FBP28 WW domain was compared to constructs in which each of the loops was replaced by CLN025. A combination of FTIR spectroscopy and laser-induced temperature-jump coupled with infrared spectroscopy was used to probe changes in the peptide backbone. The relaxation dynamics of the β-sheets and β-turn were measured independently by probing the corresponding bands assigned in the amide I region. The folding rate of the CLN025 β-hairpin is unchanged within the larger protein. Insertion of the β-hairpin into the second loop results in an overall stabilization of the WW domain and a relaxation lifetime five times faster than the parent WW domain. In both mutants, folding is initiated in the turns and the β-sheets form last. These results demonstrate that fast folding subdomains can be used to speed

the folding of more complex proteins, and that the folding dynamics of the subdomain is unchanged within the context of the larger protein.

4.2 Introduction

Protein structures are shaped by the demands of folding, stability and function in their native environment. Evolution of ultrafast folding sequences is one way nature may have balanced the need for stability of a structure against the need to maintain the flexibility necessary for it to function, or the need for fast unfolding rates for degradation and regulation.[19, 111] Exactly how nature achieves ultrafast folding is not well understood, however, at least not well enough for the rational design of fast folding sequences. Many of the fastest folding proteins that have been studied as model systems are α-helical subdomains of larger, naturally occurring proteins.[112-115] Small α-helical structures like these have a low contact order, native interactions between residues that are close in sequence, which has been correlated with fast folding.[29] These fast folding subdomains may facilitate folding of the larger proteins by acting as nucleation sites, but this has not yet been demonstrated. In contrast to helical structures, higher contact orders are found in β-sheet proteins. They exhibit slower folding rates that span a wide range.[46-48] The turns that connect multiple strands within a β-sheet fold have low contact order, however, and are thought to act as nucleation sites of folding.[72] If turn formation is rate limiting, it should be possible to speed folding by increasing the rate of turn formation. For example, optimization of the turns of Pin1 WW domain, a two-turn structure, resulted in a two orders of magnitude increase in folding of the WW domain.[77]

Here, we have explored an alternative way to increase the folding rate of a β-sheet protein, by replacing a native turn with a faster folding turn sequence. The fastest folding linear β-hairpin, the model peptide CLN025, folds within 100 ns, a similar timescale to folding of α-helical peptides.[41] It remains to be seen how this ultrafast folding β-hairpin folds in the context of a larger protein, and in turn how it impacts folding of a host system. We have addressed this question by incorporating CLN025 into WW-domains, multi-stranded β-sheet structures, that comprise two β-hairpins.[24] WW domains are good hosts for the CLN025 β-hairpin, because their small size (28-37 residues) is accessible by standard solid phase peptide synthesis and they are resistant to mutation, folding into a WW domain after modification at nearly any position.[78]

The WW domain family consists of an antiparallel and highly twisted three-stranded β-sheet structure with a small hydrophobic core and two highly conserved tryptophan residues.[73-76] Members of this family, including Pin1, FBP28 and hYAP, have been found to fold in less than 100 μs.[78,116] The fastest folding WW domain mutants are Fip35 and GTT35, derivatives of the Pin1 WW domain, which have relaxation lifetimes as fast as 14.5 μs and 3.7 μs, respectively.[77, 117] There have been extensive computational and experimental studies of WW domains, which predict folding by both two-state and three-state models.[77,78,80-87,116,118] Formin binding protein 28 (FBP28) was selected as the host WW domain for our studies, because previous work postulated that it folds through an intermediate state in which the first hairpin is highly structured.[78,86,103] Since FBP28 folds through such a polarized intermediate, it is an ideal system for exploring the effect of turn stability and folding rate on the overall mechanism. Therefore we determined the effect on the folding dynamics of inserting the optimized, fast-folding CLN025 construct into

loop 1 (residue 11-20), FBP28 1L, and loop 2 (residue 20-29), FBP28 2L, of the FBP28 structure.

We compared the folding dynamics of the mutated WW domains to the parent CLN025 and wild type FBP28 WW domain. The dynamics of WW domain formation were studied using T-jump, time-resolved IR spectroscopy. Pulsed laser excitation was used to rapidly initiate a shift in the folding equilibrium. The relaxation of the WW domain was measured by independently probing the components of the IR amide I band assigned to the β-sheet and β-turn. We found that insertion of CLN025 into the first loop has little affect on folding kinetics and stability. This is expected as previous studies report the first loop of FBP28 is structured in the intermediate-state, and thus its formation is not rate-limiting. Insertion of CLN025 into the second loop stabilizes loop 2, resulting in a global stabilization of the peptide. This mutant folds five times faster than the parent WW domain.

The folding mechanisms of both the WW domain and β-hairpin are robust. CLN025 folds at nearly the same rate inside the larger protein as it does independently, indicating that the folding mechanism is not disrupted by insertion into a larger system. Although large regional mutations were made to the native WW domain sequence, the folding mechanism that we observe is consistent with previous studies. We propose a multi-step mechanism of WW domain formation initiated in the turns. Analysis of the frequency dependence of the relaxation dynamics reveals that the turn of loop 1 forms on the nanosecond timescale. The turn of loop 2 forms second, followed by the sheet of loop 1 on the microsecond timescale. The final step in WW domain folding is formation of the second sheet. When the ultrafast folder was introduced into the second loop this

mechanism collapsed to a simple three-state model with the turns forming before the sheets. Frequency dependent IR measurements allow resolution of the folding mechanism with greater structural detail than has previously been reported. These experiments provide direct evidence for turn formation as the earliest step in the folding of a β-sheet protein, and further, that fast folding sub-domains can speed formation of larger proteins.

4.3 Experimental Section

Protein Structure Modeling. The web-service QUARK[119] was used for the prediction of the structure of a non-aggregating FBP28 WW domain mutant[99] with CLN025 β-hairpin, YYDPETGTWY,[54] inserted into one of the turns for each of the two mutants (Figure 4-1). The program used an estimated template modeling score (TM-score) to predict the quality of the protein structure prediction. A score below 0.2 corresponds to a random selection and a score above 0.5 corresponds to a high confidence in the fold.[120] The sequence with the highest TM-score when CLN025 is inserted into each loop of the WW domain was selected for peptide synthesis. The estimated TM-score of the best of the top 10 models with CLN025 inserted into loop 1, FBP28 L1, was 0.3408 ± 0.0764. The estimated Tm-score of the best of the top 10 models with CLN025 inserted into loop 2, FBP28 L2, was 0.5594 ± 0.0764.

Protein Synthesis and Purification. FBP28 1L, FBP28 2L, and wild type FBP28 were synthesized via standard 9-fluornylmethoxycarbonyl (Fmoc) based solid-phase chemistry on a Liberty1 microwave peptide-synthesizer (CEM, Matthews, NC). Fmoc-PAL-PS resin (Applied Biosystems, Foster City, CA) was used to form a peptide amide. The peptide was purified by reverse-phase chromatography (C18 column) using a

A

```
         β1        β2      β3
GATAVSEWTEYKTADGKTYYYNNRTLESTWEKPQELK  WTFBP28
GATAVSEWQEYYDPETGTWYYQNQRTQEQTWEKPQELK FBP28 1L
GATAVSEWQEYKTANGKTYYYDPETGTWYWEKPQELK  FBP28 2L
--------10--------20--------30
```

B

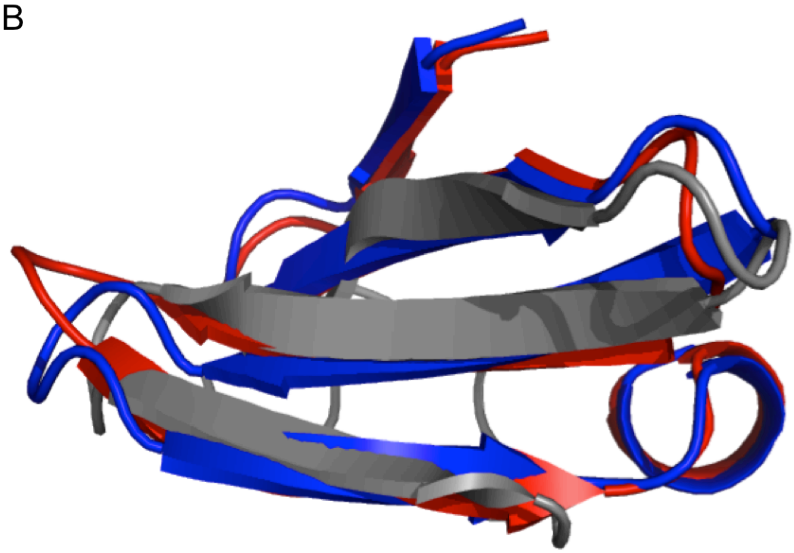


Figure 4-1. Prediction of FBP28 1L and FBP28 2L structures. (A) Sequence alignment of wild type FBP28 protein with synthetic constructs containing the CLN025 sequence (in blue). The three β-strands are labeled above and colored red in the FBP28 sequence. (B) The three-dimensional models constructed using QUARK coordinates aligned with the FBP28 structure (PDB entry: 1E0L). The wild type FBP28 is shown in grey, FBP28 1L is shown in red, and FBP28 2L is shown in blue. The figure was prepared using PyMOL (www.pymol.org).

water/acetonitrile gradient with 0.1% trifluoroacetate (TFA) as the counter-ion. TFA interferes in the Amide-I IR measurements at 1672 cm^{-1}, so we remove it by anion exchange. The peptide was lyophilized and dissolved in a 2 mM HCl solution to allow exchange of the TFA counter-ion for chloride.[90] The identity of the peptide was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. The peptide was lyophilized and dissolved in D_2O to allow deuterium-hydrogen exchange of the amide protons. The peptide was lyophilized a second time and resuspended in D_2O buffer with 20 mM potassium phosphate buffer at pD* 7.0 (pD* refers to the uncorrected pH meter reading). Sample concentrations of 0.5-1.0 mM were prepared for both IR and fluorescence experiments.

CD Spectroscopy. CD spectra and CD melting curves were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Jasco temperature controller module (Jasco, Inc, Easton, MD). Peptides were dissolved at 50 μM in 20 mM potassium phosphate buffer at pH 7.0. All measurements were obtained using a 1-mm pathlength cell. Wavelength scans were recorded over the range of 260 nm to 190 nm with an average of 3 repeats. A bandwidth of 2 nm and scan rate of 50 nm/min were used for spectral acquisition. Thermal unfolding experiments were performed by monitoring the signal at 226 nm from 5 °C to 90 °C using a 0.1 °C interval and scan rate of 30 °C/hr. During the thermal unfolding experiment a full wavelength scan was obtained every 5 °C after a 60 second delay. The buffer and protein concentrations were the same as used in the wavelength scan experiment.

FTIR Spectroscopy. The equilibrium melting behavior was monitored on a Varian Excalibur 3100 FTIR spectrometer (Varian Inc., Palo Alto, USA) using a temperature

controlled IR cell. The IR cell consists of two CaF_2 windows stacked and separated by a 100 μm Teflon spacer split into two compartments, a sample and a reference. The same cells are used for equilibrium FTIR and T-jump experiments. No aggregation was observed in the infrared at reported concentrations. All spectra shown at a specific temperature are constructed by subtracting the spectrum of reference buffer solution without protein from sample solution with protein. The temperature-dependent difference spectra were then generated by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures. The second derivative spectra were computed in IGOR PRO after smoothing the data with a sixth order binomial algorithm to remove any residual water vapor (WaveMetrics, Lake Oswego, OR).

Time Resolved Temperature Jump (T-jump) Relaxation Measurements. The IR T-jump apparatus has been described previously.[30] Pulsed laser excitation is used to rapidly perturb the folding equilibrium on a timescale faster than the molecular dynamics of interest. Time resolved infrared is then use to probe the reaction. A Q-switched DCR-4 Nd:YAG laser (Spectra Physics, Mountainview, CA) fundamental at 1064 nm is Raman shifted (one stokes shift in 200 psi H_2 gas) to produce a 10 ns pulse at 2 μm. The magnitude of the T-jump is calculated using the change in reference absorbance with temperature. The T-jump reference is taken from D_2O buffer with 20 mM potassium phosphate buffer at pD* 7.0 at the same temperature and frequency as the sample. Absorbance changes at the reference frequency are due only to changes in D_2O absorbance, which is used as an internal thermometer.[30]

The change in signal induced by the T-jump is probed in real time by a continuous laser with a frequency in the amide I' band of the IR. The mid-IR probe beam is

generated by a continuous wave quantum cascade laser (Daylight Solutions Inc., San Diego, CA) with a tunable output range of 1570-1730 cm^{-1}. The transient transmission of the probe beam through the sample is measured using a fast, 100 MHz, photovoltaic MCT IR detector/preamplifier (Kolmar Technologies, Newburyport, MA). Transient signals are digitized and signal averaged (1000 shots) using a Tektronics digitizer (7612D, Beaverton, OR). Instrument control and data collection are controlled using a LabVIEW computer program.

Analysis of Kinetics Data. The peptide relaxation kinetics must be deconvolved from the observed kinetics. Accurate deconvolution is possible as the instrument response is determined from the reference measurement under the exact conditions of the sample measurements. In order to minimize detector artifacts, the reference is scaled prior to subtraction from the sample. The decay function is a triple exponential decay with the formula:

$$A = A_0 + \ldots + A_n \exp\left(\frac{-(x - x_o)}{\tau_n}\right) \qquad (4\text{-}1),$$

where A_0 is an offset, n is the number of exponentials to fit, A_n is a preexponential factor, τ_n is the relaxation lifetime of the sample and x_0 is the time offset. In order to best fit the data, the minimum number of exponentials with unique relaxation lifetimes was selected. The data are fit over the interval from 90 ns to an order of magnitude outside the slowest exponential. In the cases where a fast phase is not reported the fit starts at 400 ns. The data analysis was performed in IGOR PRO (WaveMetrics, Lake Oswego, OR).

Results and Discussion

4.4 QUARK Prediction of WW domain Structure. The web-service QUARK[119] was used to predict the Formin binding protein 28 (FBP28) WW domain structure with the fast-folding CLN025 β-hairpin inserted into each of the loops (Figure 4-1). Glutamine and asparagine mutations were made to the wild type FBP28 sequence to eliminate aggregation.[99] There is good agreement between the tertiary structure prediction of the mutated sequences and the wild type WW domain (Figure 4-1B). The locations of each of the sheets, β1, β2, and β3, align well with the NMR structure of the FBP28 WW domain (PDB ID 1E0L). In both models QUARK predicts two additional secondary structure elements, an N-terminal β-sheet and a C-terminal α-helix. These small differences at the termini of the peptide do not affect the arrangement of the core structure. There is no experimental evidence by either CD or IR spectroscopy that these additional structural elements are formed, meaning they are likely artifacts of the QUARK prediction.

4.5 Far-UV CD Spectroscopy. The FBP28 1L and FBP28 2L peptides were constructed to test the affect of CLN025 insertion into each loop of FBP28 on WW domain folding. Far-UV CD is a good indicator of peptide secondary structure. The typical β-sheet signal has a negative peak at ~218 nm and a positive peak at ~195 nm.[91] Folded WW domains do not exhibit this behavior; instead they have a negative peak at ~202 nm and a positive peak at ~230 nm.[92, 93] Small variations in the position and intensities of these peaks have been observed across the WW domain family.[93] The peak at ~202 nm resembles the random coil peak usually found at 200 nm. Disorder in the N- and C- termini of the folded structure of WW domains is thought to contribute to the

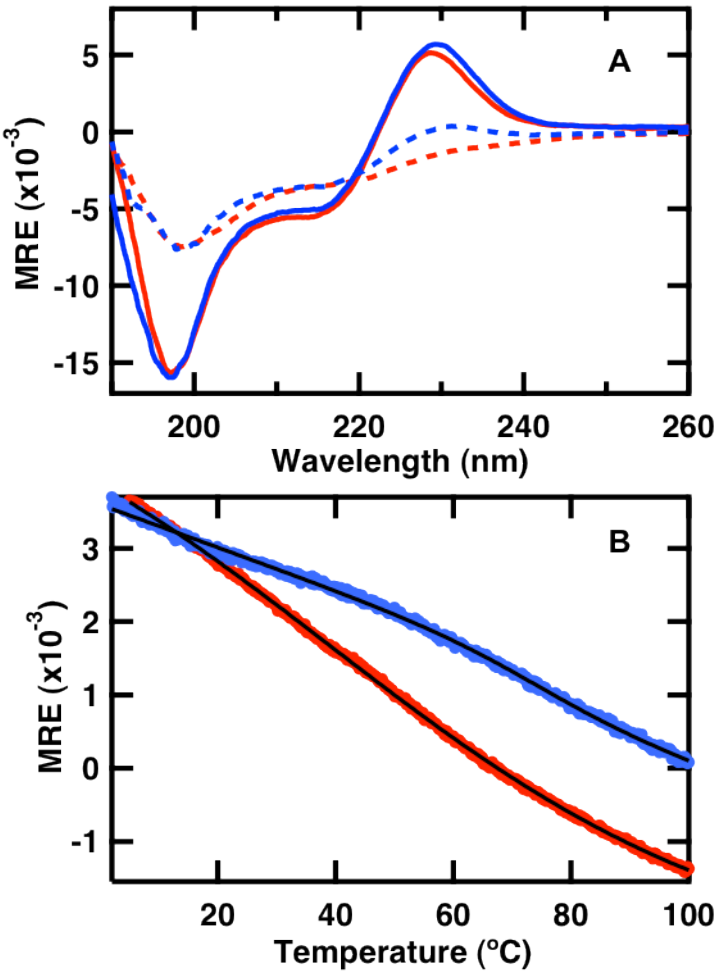


Figure 4-2. (A) Far-UV CD spectra of 50 μM solutions of the FBP28 1L (red) and FBP28 2L (blue) WW domains in 20 mM potassium phosphate buffer (pH 7.0) acquired at 2 °C (____) and 95 °C (----) during the course of a thermal denaturation in a 0.1 cm pathlength cell. (B) Thermal denaturation of the WW domains monitored by CD at 226 nm. The continuous line represents the best fit to the data to equation (4-2).

negative peak in the CD spectrum.[92, 94-96] The peak at ~230 nm arises from the presence of ordered aromatic amino acids.[91, 92, 116] Analysis of the CD spectrum of FBP28 1L and 2L

reveals both a positive peak at 226 nm and a negative peak at 197 nm (Figure 4-2A), consistent with a WW domain topology and the structure predicted by QUARK.

Thermal denaturation was monitored by recording the absorbance change at 226 nm with temperature (Figure 4-2B). The WW domains exhibit the typical heat-induced unfolding behavior with a loss of intensity at 226 nm and a shift of the minimum at 197 nm to 200 nm, corresponding to a change in the secondary structure from WW domain to random coil (Figure 4-2A).[92] The melting curves (Figure 4-2B) were fit to an apparent two-state equilibrium model:

$$A = A_i + mT + \frac{A_f}{1 + \exp\left(\left(T_m - T\right)/\Delta T\right)} \qquad (4\text{-}2),$$

where A_i and A_f are the extrapolated absorbance values at the two endpoints of the transition taking into account a linear baseline m, T_m is the transition midpoint and ΔT is the overall temperature range of the transition. The transition for FBP28 1L was too broad to obtain a reliable fit. FBP28 2L has a greater thermodynamic stability than FBP28 1L; at the highest temperature (Figure 4-2A) of the melt of FBP28 2L, there is still some intensity in the positive band at 230 nm, indicating that the peptide has not completely unfolded. The melting temperature of FBP28 2L is 74.4 ± 0.4°C. This is 23 °C higher than the melting temperature of FBP28, and closer to the 69.9 °C melting temperature of the CLN025 β-hairpin.[54, 99] The transition of FBP28 2L is also much sharper (smaller ΔT) than the transition of FBP28 1L, revealing an increased cooperativity in the melt of FBP28 2L.[50] This is evidence that insertion of CLN025 into the second loop of FBP28 stabilizes the WW domain.

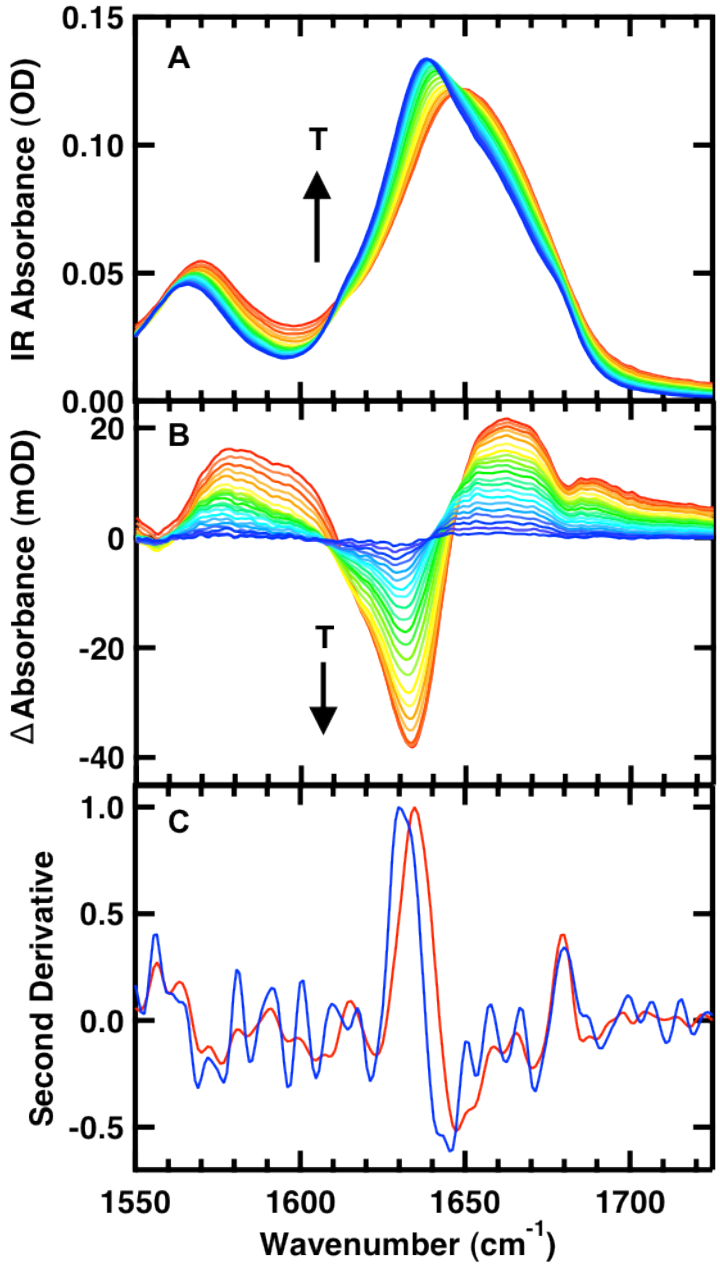


Figure 4-3. Temperature dependent FTIR spectra of 1 mM FBP28 2L in 20 mM potassium phosphate buffer (pH 7). (A) Absorbance spectra in the Amide I' region; the temperature of the individual traces varies from 5 to 85 °C in 5 °C intervals. (B) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher

temperatures. (C) Normalized second derivative of FTIR difference spectrum at 33 (blue) and 67 °C (red).

4.6 FTIR Spectroscopy. The temperature-induced unfolding of the two mutants was studied over the range from 5 to 95 °C in 5 °C intervals using FTIR spectroscopy to monitor the amide I' region. An example of the temperature dependent absorption spectra of the amide I' spectral region (amide I region of peptides in D_2O) of FBP28 2L is shown in Figure 4-3A (similar data for FBP28 1L is shown in Appendix Figure 4A-1). The amide I' absorbance arises from C=O stretching vibrations of the polypeptide backbone carbonyls, and is an established indicator of secondary structure.[62-64] This relatively broad band contains contributions from the entire polypeptide backbone, which in the case of the WW domain includes β-sheet, β-turn, and random coil structure. The changes with temperature are highlighted by the difference spectra for each peptide (Figure 4-3B). The difference spectra are generated by subtracting the lowest temperature spectrum from each absorbance spectrum at higher temperature. Negative peaks correspond to specific structures or interactions present in the folded state, and positive peaks correspond to new interactions with solvent in the unfolded state. The individual peaks are more easily distinguished in the second derivative of the FTIR spectra at the lowest temperature (Appendix Figure 4A-2).

At low temperature there are three major peaks that are characteristic of specific secondary structures, centered at 1614, 1634, and 1679 cm^{-1}. In the difference spectra the peak at 1614 appears as a shoulder on the dominant negative component at 1634 cm^{-1}. The negative peak at 1679 cm^{-1} overlaps with the broad positive spectral feature at 1660

cm^{-1}, a characteristic peak due to the disordered polypeptide at high temperatures.[30, 43, 65] The Gai group previously observed these peaks in the Pin1 WW domain.[100] A peak at 1614 cm^{-1} was observed in CLN025 and attributed to a tertiary amide interaction between the glutamic acid backbone in the turn and the aspartic acid side chain.[41, 66] In the FBP28 WW domain there is a tertiary amide interaction between the backbone of lysine and side chain of threonine in the first turn and there are no native tertiary amide interactions in the second turn.[74, 101] We assign the peak at 1614 cm^{-1} to tertiary amide interactions in the turns of the WW domain. In FBP28 1L the peak at 1614 cm^{-1} is assigned to tertiary amide interactions in the first loop, found in the turn of CLN025. In FBP28 2L the peak is assigned to tertiary amide interactions in the turns of both loops, stemming from the FBP28 WW domain in loop 1 and CLN025 in loop 2. IR bands at 1634 and 1681 cm^{-1} are well-established markers of antiparallel β-sheets.[50] The peak at 1634 cm^{-1} arises from inward directed carbonyl groups in the sheet and the peak at 1681 cm^{-1} arises from solvent exposed carbonyl bonds in the sheet. The FBP28 amide I' bands at 1634 cm^{-1} and 1679 cm^{-1} are consistent with these characteristic β-sheet markers.

Analysis of the second derivative of the difference spectra with temperature (Figure 4-3C) highlights the temperature dependent contributions to the 1634 cm^{-1} difference peak. Below 46 °C the second derivative peak maximum is at lower frequency, 1629 cm^{-1}. As the temperature increases, this peak shifts to a higher frequency, 1635 cm^{-1} (slightly higher than the peak observed in the difference spectrum, because the second derivative represents the actual peak maximum). A similar phenomenon was observed by the Keiderling group in their extensive studies of multi-stranded β-sheet structures.[102] They assign the main maximum to out-of-phase amide I vibrations of neighboring amides in

the β-sheet. As the number of strands was increased, the peak maximum decreased in frequency. This is due to interstrand coupling as the number of β-sheets increases. The efficiency of the interstrand coupling depends on the orientation of the dipole moments of the individual carbonyl oscillators. The splitting is less in WW domains than other extended β-sheet structures, because the twist of the sheets misaligns the carbonyl oscillators so that they do not couple as efficiently as in more planar structures, causing only a small change in frequency with strand number. We attribute the second derivative 1635 cm^{-1} feature to C=O bonds in the β-sheet of a single hairpin. As the extended three β-strand WW domain structure forms, there is a shift of the peak maximum to lower frequency, 1629 cm^{-1}. In the difference spectra (Figure 4-3B), as the peptide is melted we observe the melt of the first strand of the WW domain, loss of the 1629 cm^{-1} band, followed by loss of the band at 1635 cm^{-1} due to the second strand. We conclude that the second loop melts first, since the first loop has been previously identified as the more stable of the two loops.

The normalized melting curves for both proteins derived from the temperature dependent IR absorbance at 1634 cm^{-1} are shown in Figure 4-4. The data were normalized after being fit to an apparent two-state equilibrium model using equation 4-1. The melting temperature of FBP28 2L measured at 1634 cm^{-1} is 65.3 ± 0.8°C. The melting temperature derived from fits to other IR frequencies agreed within the error of the fit. The melting temperature derived from the IR melt of FBP28 2L is similar to that reported for CLN025, 69.9 °C.[54] Thermodynamic parameters were derived from a van't Hoff analysis (inset of Figure 4-4) of the FTIR data at 1634 cm^{-1}. The free energy of unfolding of these peptides over the temperature range studied is linear. The

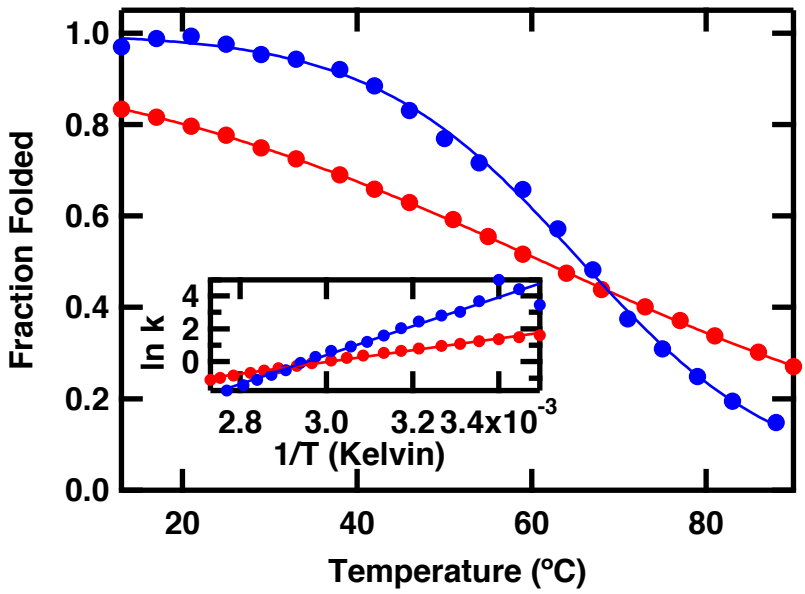


Figure 4-4. FTIR melt curves for the FBP28 1L (red) and FBP28 2L (blue) WW domains obtained by plotting the change in IR difference spectra at 1634 cm^{-1} versus temperature. The data are fit to an apparent 2-state model (Equation 4-2) and then normalized. Inset: Van't Hoff analyses using equilibrium constants generated from the melt curves.

thermodynamic parameters observed in FBP28 2L ΔH_f, -73 ± 4 kJ/mol, and ΔS_f, -216 ± 11 J/mol•K, are similar to the ΔH_f, -62.8 kJ/mol, and ΔS_f, -182.8 J/mol•K reported for the CLN025 hairpin.[41] Since the thermodynamic parameters of FBP28 2L mimic those of the isolated hairpin, we conclude that insertion of CLN025 into the second turn of FBP28 stabilizes the WW domain.

4.7 Temperature-Jump Relaxation Kinetics. The relaxation kinetics of the folding/unfolding transition following a laser induced temperature jump were probed using time resolved infrared spectroscopy of the amide I' frequency of the tertiary amide linkage of the turns (1619 cm^{-1}) and the β-sheet of the first loop (1633 cm^{-1}) and β-sheets

of the fully formed WW-domain (1629 cm^{-1}). Jumps were performed slightly off peak center to maximize the signal. Figure 4-5 displays the relaxation kinetics for the peptides following a jump from 30 to 45 °C (Figure 4-5A) and from 50 to 65 °C (Figure 4-5B). These jumps are to the midpoint of the transition of FBP28 1L and FBP28 2L, respectively. The relaxation transient is fit by a triple exponential at low temperature and a single or double exponential at high temperature. The complete relaxation parameters observed for the peptides are reported in the appendix (Appendix Table 4A-1, 4A-2, 4A-3). Both peptides exhibit three relaxation lifetimes (Figure 4-6). The first two, a fast 100 ns phase (τ_1) and a μs (τ_2) phase, are relatively temperature independent. The third relaxation lifetime (τ_3) is highly temperature dependent. The observed relaxation of τ_3 spans 100's of microseconds through a few μs. As the temperature increases, τ_2 becomes indistinguishable from τ_3.

At jumps to low final temperatures there is little difference between the observed kinetics and relative intensities of the relaxation lifetimes of FBP28 1L and 2L (Figure 4-5A). The absorbance change associated with each of the three phases is of approximately equal intensity (Appendix Table 4A-1, 4A-2, 4A-3). At jumps to higher temperatures FBP28 1L and FBP28 2L exhibit different behaviors (Figure 4-5B). Most of the absorbance change in FBP28 1L is observed in the slow phase whereas most of the absorbance change in FBP28 2L is in the fast phase, τ_1. An Arrhenius plot (Figure 4-6) of the observed kinetics reveals that at low temperature there is little difference between the kinetics of the three phases of FBP28 1L and FBP28 2L. At higher final jump temperatures, although not all the phases are observed, the kinetics that are observed follow the same trend.

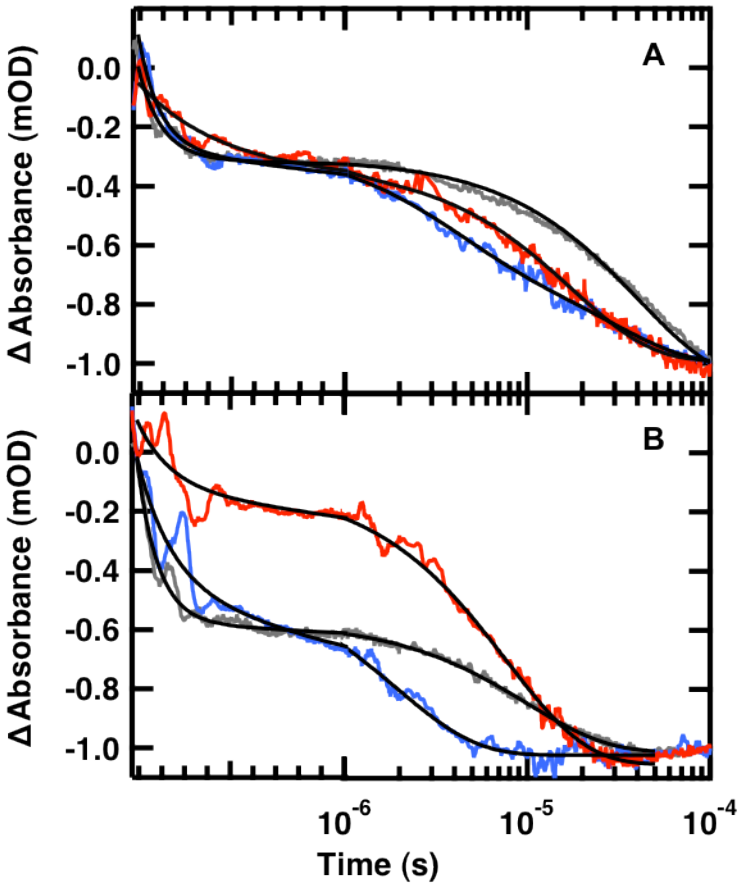


Figure 4-5. Representative IR T-jump relaxation kinetics of FBP28 1L (red), FBP28 2L (blue) and wild type FBP28 (grey) monitored in the amide I' spectral region at 1633 cm^{-1} following a T-jump from (A) 30 to 45 °C and (B) 50 to 65 °C. A triple exponential fit, equation (4-1), is overlaid on each kinetic trace on top and a double exponential fit is overlaid on each kinetic trace on the bottom (black solid line). The data are normalized at $1x10^{-4}$ for comparison.

Frequency dependent T-jump measurements (Figure 4-7) reveal differences in the kinetics of the turns (1619 cm^{-1}), the β-sheet of a single hairpin (1633 cm^{-1}) and β-sheet of the complete WW domain (1629 cm^{-1}). There are multiple exponentials observed at

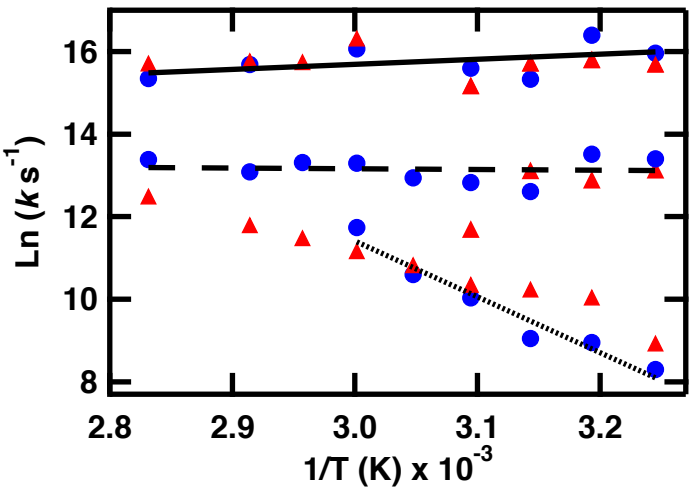


Figure 4-6. Arrhenius plot showing the temperature dependence of the folding region at 1633 cm^{-1}. The values of T used for the (1/T) axis are the final temperatures reached during the jump. *k* is the value obtained from a fit (see text) of the T-jump transient of FBP28 1L (▲) and FBP28 2L (●). Lines are a result of fitting τ_1 (solid line), τ_2 (dashed line), and τ_3 (dotted line) of FBP28 2L.

each of these probe frequencies due to overlap between the bands. It is also likely that the spectral response is coupled such that formation of one structure affects the others. The following discussion focuses on FBP28 1L, which has the ultrafast folding CLN025 sequence in the first loop. The 1619 cm^{-1} band in this case arises exclusively from the first turn, as there are no tertiary amide interactions in the second turn. The kinetics monitored at 1619 cm^{-1} exhibits the largest amplitude of the fast (~100 ns) phase (Table 4-1). Based on this observation we assign the fast phase to turn formation in loop 1 and the intermediate (8 μs) phase to sheet formation. This supports the hypothesis that folding is initiated in the turn, where there is lower contact order.[29] The kinetics at 1633 cm^{-1} is

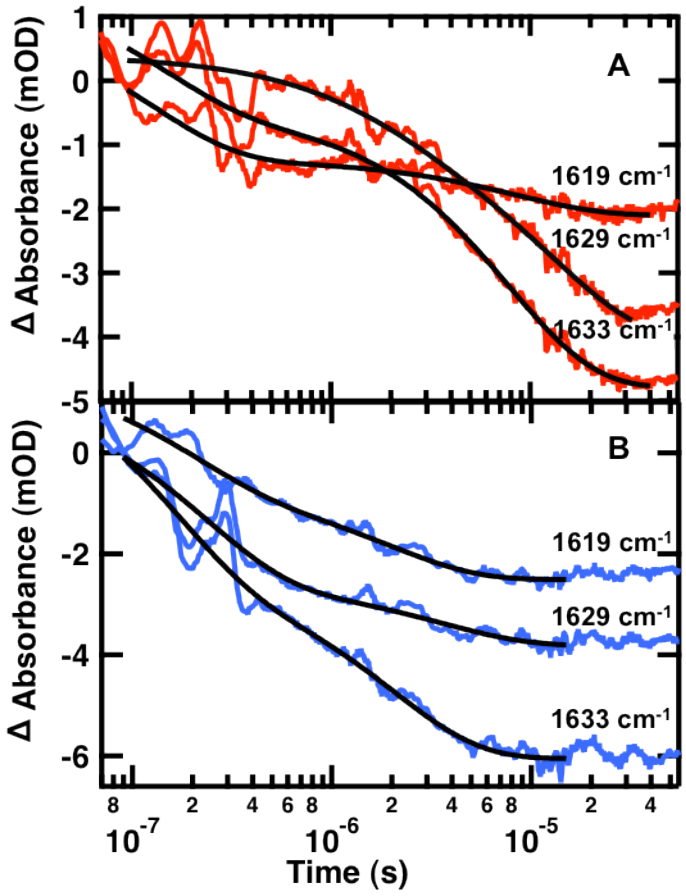


Figure 4-7. Representative IR T-jump relaxation kinetics of FBP28 1L (A) and FBP28 2L (B) monitored in the amide I' spectral region at 1619, 1629, and 1633 cm^{-1} following a T-jump from 50 to 65 °C. A double exponential fit is overlaid on each kinetic trace (black solid line).

sensitive to the β-sheet of a single hairpin. The observed kinetics is similar to that at 1619 cm^{-1}, which probes the turn of the first loop, but with more intensity in the intermediate phase than the fast phase. The 1633 cm^{-1} band in FBP28 1L was assigned to the β-sheet of the first loop based on its higher stability in the temperature dependent FTIR measurements. Thus we assign the fast (~100 ns) phase to the formation of the turn in

Table 4-1: Relaxation Kinetics Following a Jump from 50 to 65 °C

λ	FBP28 1L				FBP28 2L			
(cm^{-1})	A_1 (%)	τ_1 (ns)	A_2 (%)	τ_2 (μs)	A_1 (%)	τ_1 (ns)	A_2 (%)	τ_2 (μs)
1619	65±2	105±6	35±1	8.6±0.5	45±3	185±15	55±2	1.7±0.8
1629	28±8	2300±400	72±1	17.2±0.4	67±2	213±10	33±1	3.1±0.2
1633	20±1	151±14	80±1	7.9±0.1	42±2	154±11	58±1	2.1±0.1

loop 1 and the intermediate (8 μs) phase to sheet formation in loop 1. This conclusion is consistent with previous studies that postulated FBP28 folds through an intermediate in which the first loop is fully formed. Replacing the first loop with the ultrafast folding CLN025 hairpin should not perturb this mechanism.

The probe at 1629 cm^{-1} is sensitive to interstrand coupling across all three sheets of the WW domain. Since formation of the second loop is thought to be the rate-limiting step for WW domain formation, the kinetics at 1629 cm^{-1} should probe this final folding step. As expected, most of the 1629 cm^{-1} intensity is in the slowest (17 μs) phase, which we assign to β-sheet formation in loop 2. We postulate that the 2.3 μs phase observed at 1629 cm^{-1} is the turn of loop 2; however, we do not have a direct probe of the second turn of FBP28 1L. Similar to previously studied WW domains, we propose a mechanism of FBP28 1L WW domain formation initiated in the turn of loop 1 (Figure 4-8). The turn of loop 2 forms second, followed by the sheet of loop 1. The final step in WW domain formation is the completion of the second loop.

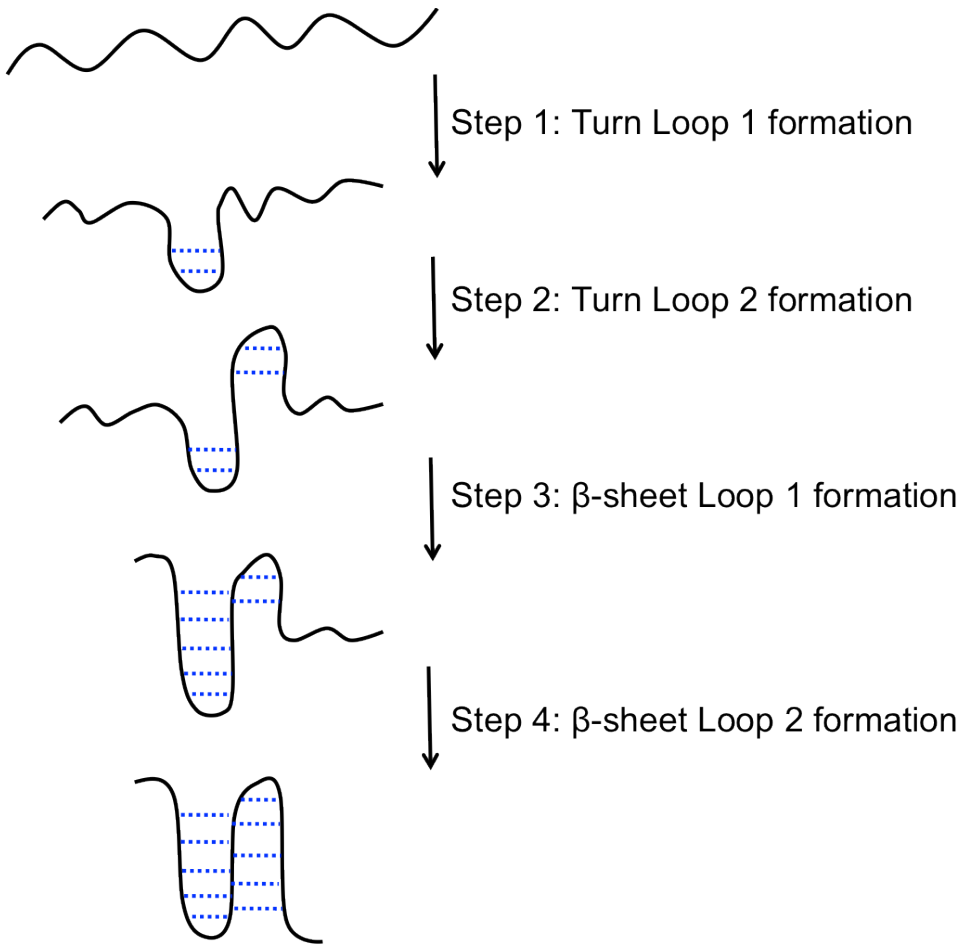


Figure 4-8. Proposed mechanism of WW Domain folding

Using this model we can interpret the three phases observed in the Arrhenius analysis (Figure 4-6). The folding mechanism of FBP28 1L includes four steps (Figure 4-8), one more than was observed in the Arrhenius analysis at 1633 cm^{-1}, a probe of loop 1. Because the IR band at 1633 cm^{-1} is relatively broad and overlaps with the band at 1619 and 1629 cm^{-1}, there may be a relatively smaller contribution from kinetics associated with each of these wavelengths. The ns and μs temperature independent fast phases are assigned to the formation of the turn and sheet of loop 1, respectively. The third phase that is highly temperature dependent, is assigned to formation of the β-sheet of loop 2. At

high temperatures, the dynamics of the sheet of loop 2 and sheet of loop 1 become indistinguishable. Analysis of frequency dependent measurements and relative intensities of the absorbance changes give strong evidence for a fourth phase, formation of turn 2, in between the formation of the turn and sheet of loop 1 (Table 4-1). This turn forms on a similar timescale to the sheet of loop 1, the dominant contribution monitored at 1633 cm^{-1}, so it is not possible to distinguish a fourth phase in the Arrhenius analysis (Figure 6).

The following discussion focuses on the kinetics of FBP28 2L, which has the ultrafast folding turn inserted in the second loop. Whereas speeding the formation of the first loop of FBP28 has little affect on the observed dynamics and folding mechanism, speeding the formation of the second loop speeds the overall folding dynamics and simplifies the folding mechanism (Table 4-1). In FBP28 2L, measurements at 1619 cm^{-1} include contributions from both turns, because there are tertiary amide interactions found in both of the turns. The fast kinetics is again assigned to the turns. It is not possible to determine which turn is forming first as it is not possible to independently probe each turn. The slower phases are assigned to β-sheet formation. When the kinetics is probed at either the frequency associated with the β-sheet of a single hairpin or that of the complete WW domain (1633 and 1629 cm^{-1}, respectively), there is only a small difference between the observed slow rates. Speeding the folding of the second loop greatly speeds up the rate-limiting step, formation of the β-sheet of the second loop, simplifying the folding model. The folding of FBP28 2L WW domain is initiated in the turns followed by zipping of the β-sheets.

Previous fluorescence studies of FBP28 revealed biexponential kinetics below 65 °C with relaxation lifetimes of 15 μs and >900 μs.[84] Above 65 °C single exponential kinetics

were reported, with a relaxation lifetime of <15 μs.[78] Infrared measurements offer a level of detail not available from the fluorescence experiments. IR measurements of the wild type FBP28 (Figure 4-5) and mutants reveal fast nanosecond and microsecond phases that have not been previously observed in this sequence. A jump to a final temperature of 65 °C shows that mutation of the second loop results in an order of magnitude increase in the observed relaxation lifetimes (Table 4-1). The same experiment with the mutation to the first loop results in no change in the folding dynamics. Insertion of the CLN025 hairpin into loop 2 speeds its folding, allowing the WW domain to form faster. The relaxation lifetime of the fastest folding WW domain, GTT35, is 3.7 μs at a temperature for which the equilibrium constant for folding is ~2.[77] Comparing the relaxation lifetime of the FBP28 2L mutant at the same stability, a final jump temperature of 60 °C, the relaxation lifetime is <5 μs, comparable to the GTT35 WW domain. GTT35 is a mutant of the Pin1 WW domain, which resulted from optimization of both of the turns in the WW domain. Our experiments show that large regional mutations is an effective approach to increase the folding rate of a larger system.

Protein mutations have mixed affects on the mechanism of protein folding. SH3 domains were found to have a discrete order to folding. Computational studies of src, Fyn and α-spectrin SH3 domains showed that while there is local variation in the transition state ensembles of each, they share a topology characteristic of the SH3 domain.[121] Experiments on SH3 domain where mutants included lengthening of the n-src by 10 residues and an introduction of disulfide crosslinks all exclusively affected the folding rates.[122] Similarly the folding mechanism after mutation of the first loop of the FBP28 WW domain is consistent with previous studies. CLN025 folds at nearly the same rate

inside the larger protein as it does independently, indicating that its folding mechanism is not disrupted by insertion into a larger system. Studies of protein G and protein L, which share a similar topology, show different transition state ensembles where the hairpin with the lowest free energy forms first.[123] The overall order of folding was switched in protein G, by changing the stability of each of the turns.[124] The folding of the slower turn was sped by selecting a low energy sequence with termini that superimposed on the terminal residues of the hairpin. However, in order to change the folding mechanism a hydrogen bond in the faster turn was disrupted. The folding of the protein G mutant without this point mutation in the turn was five times faster. While this shows that it is possible to change the order of folding, it came at the cost of the folding rate. We observe that the folding rate is limited by the folding of the slower turn. Our measurements of FBP28 2L (Table 4-1) show that the insertion of the fast folding β-hairpin into the second turn speeds not only the formation of the second turn, but also of the sheets of both loop 1 and loop 2. Increasing the rate of folding of the turns is one way that nature may create faster folding β-sheet proteins. We have shown that small fast folding single domain β-hairpin models can be used to speed the folding of a larger protein without exhaustively searching for low energy sequences.

4.8 Conclusion

Insertion of the CLN025 β-hairpin into each loop of the FBP28 WW domain reveals that the folding dynamics of the ultrafast folding hairpin is the same within the context of a larger protein. CLN025 folds in ~100 ns both independently[41] and inside the WW domain. Furthermore, the ultrafast folding hairpin accelerates the overall folding of the

WW domain by an order of magnitude when placed in loop 2, since the formation of this loop clearly limits the overall rate of folding. Previous studies have focused on speeding the folding of larger systems by point mutations or sequence homology based on the energy of the sequence.[77, 124] While these methods have proven effective, we have demonstrated an alternative method that introduces a fast folding subdomain to nucleate the larger structure. Since the folding of the CLN025 domain is unchanged in the context of the larger WW domain, it is a robust nucleation structure that may be generally applicable to speeding the folding of other β-sheet systems.

We propose a model to describe WW domain folding initiated in the turns, similar to that proposed for other WW domains.[78, 86, 103] Previous studies using fluorescence T-jump have been unable to observe the level of detail available from infrared spectroscopy. We are able to observe subtle differences in the timescale of formation of each of the turns and sheets of the loops in the WW domain. We observe that the turn of both loop 1 and loop 2 form prior to the β-sheets. This highlights the importance of the turn in folding of β-proteins. The folding mechanism of the WW domain is robust and is not disrupted by mutation of the loops. We observe that the speed of folding is limited by formation of the second turn. When the fast folding β-hairpin is inserted into the second turn, the folding of the sheets of both loops is increased. Thus, by applying our understanding of the basic principles of protein folding and small subdomains and models like the CLN025 β-hairpin we can speed formation of larger multi-domain systems.

4.9 Appendix

1. Temperature dependent FTIR spectra of FBP28 1L were recorded to confirm that the peptide was folded. Temperature dependent FTIR of FBP28 1L are reported in Figure 4A-1.

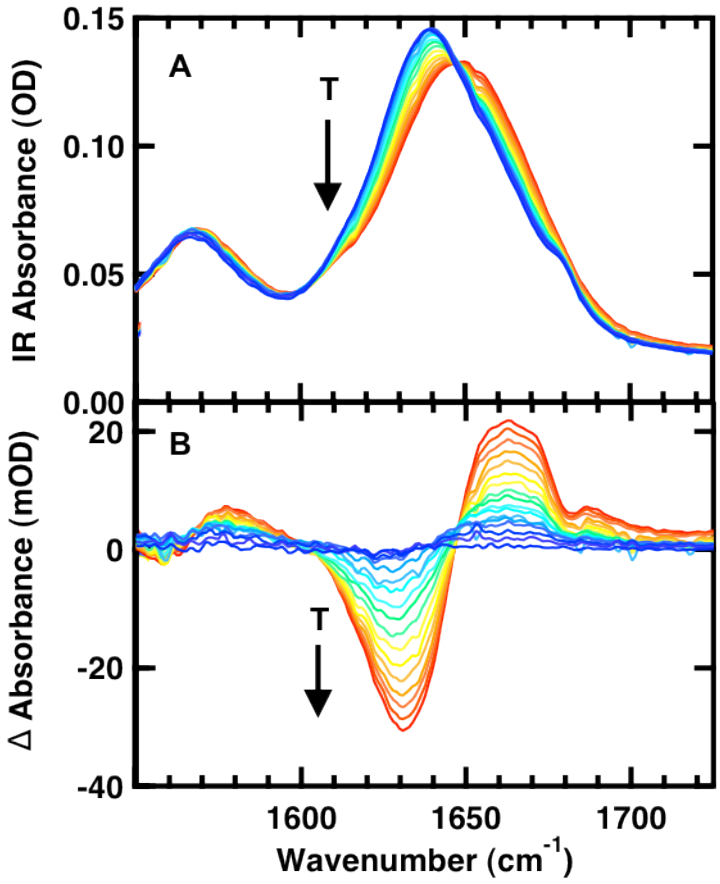


Figure 4A-1. Temperature dependent FTIR spectra of 1 mM FBP28 1L in 20 mM potassium phosphate buffer (pH 7). (A) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 5 to 85 °C in 5 °C intervals. (B) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher temperatures.

2. Second derivative spectra of FBP28 1L and 2L absorbance data at 5 °C and FBP28 1L difference data at 33 and 67 °C are taken to determine the peaks in absorbance data. The second derivatives were obtained after smoothing the data with a third order (sixth for difference data), binomial algorithm to remove any residual water vapor lines. The data analysis was performed in IGOR PRO (WaveMetrics, Lake Oswego, OR). The individual peaks in the absorbance spectra are more easily distinguished in the second derivative spectra. The second derivative of the absorbance data (Figure 4A-2) shows three peaks corresponding to secondary structure, centered at 1614, 1638, and 1679 cm^{-1}. The second derivatives of the difference spectra at 33 and 67 °C highlight the peaks that are changing with temperature (Figure 4A-3). At low temperature a peak at 1629 cm^{-1} dominates and at high temperature a peak at 1634 cm^{-1} dominates. The data is normalized to the maximum near 1634 cm^{-1} for comparison.

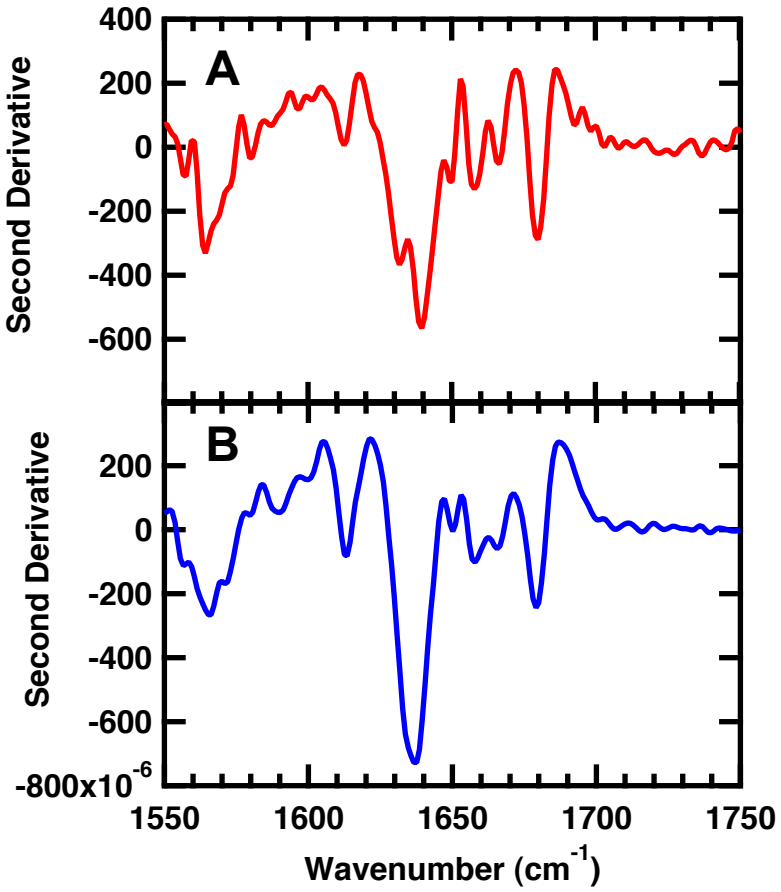


Figure 4A-2. Second derivative of FTIR absorbance spectrum of FBP28 1L (A) and FBP28 2L (B) at 5 °C.

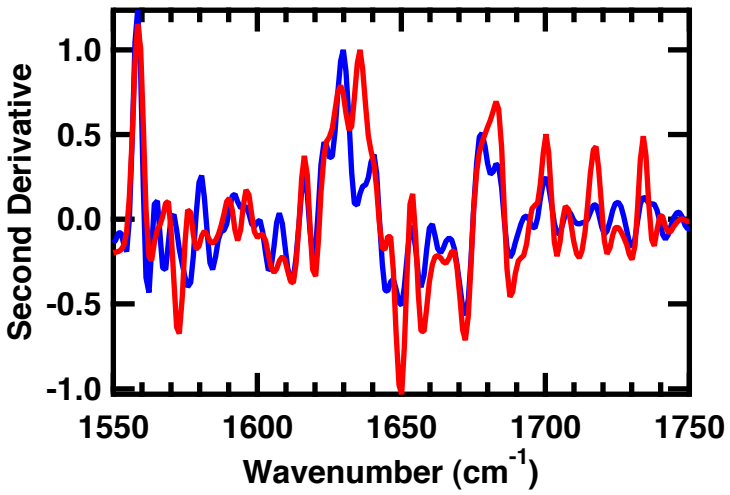


Figure 4A-3. Second derivative of FTIR difference spectrum of FBP28 1L at 33 (blue) and 67 (red) °C.

3. Temperature dependent infrared T-jump measurements were collected to determine the dynamics of the systems. Wavelength dependent IR T-jump measurements were collected at 1619 (Table 4A-1), 1629 (Table 4A-2) and 1634 cm^{-1} (Table 4A-3). 15 °C jumps were performed with an initial temperature from 15 to 60 °C. Data reported includes the observed magnitude and relaxation lifetime of each event.

Table 4A-1: Relaxation Kinetics Probed in the Turn of Loop 1 (1619 cm^{-1}) of FBP28 1L and FBP28 2L

FBP28 1L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30						
20-35						
25-40						
30-45						
35-50						
40-55	1.55±0.03		1.40±0.01	306±11		10.5±0.4
45-60			1.12±0.01			12.7±0.3
50-65	1.22±0.05		0.96±0.04	117±8		11.0±1.3
60-75	1.03±0.11		2.12±0.03	98±18		6.3±0.3
FBP28 2L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30						
20-35						
25-40						
30-45						
35-50						
40-55						
45-60	1.38±0.05	1.44±0.04		131±10	1.83±0.8	
50-65	1.39±0.06	1.70±0.05		185±15	1.7±0.8	
60-75	3.17±0.05			150±4		

Table 4A-2: Relaxation Kinetics Probed in the Sheet of the WW Domain (1629 cm^{-1}) of FBP28 1L and FBP28 2L

FBP28 1L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	0.78±0.02	0.96±0.02	0.69±0.01	140±8	2.03±0.06	126±3
20-35	1.06±0.02	1.16±0.01	0.97±0.01	134±5	2.58±0.07	40.8±0.9
25-40	1.02±0.02	1.09±0.01	0.75±0.01	125±5	2.23±0.06	37.6±0.6
30-45	0.96±0.02		0.76±0.01	160±7		28.1±0.9
35-50			2.15±0.01			14.3±0.2
40-55	0.50±0.03		2.45±0.01	158±15		16.3±0.2
45-60			3.23±0.01			9.3±0.1
50-65		1.17±0.18	3.31±0.09		2.2±0.3	17.4±0.3
60-75	1.31±0.07		3.18±0.07	122±11		20.8±1.0
FBP28 2L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30						
20-35						
25-40	1.52±0.02	1.11±0.02	0.72±0.01	260±8	2.9±0.1	130±5
30-45	1.50±0.03	1.57±0.03	0.60±0.02	203±8	2.48±0.09	43.6±3
35-50	0.70±0.06	1.79±0.10	0.26±0.09	102±17	1.8±0.1	11.3±7
40-55	1.72±0.06	2.49±0.05		157±12	1.74±0.05	
45-60		2.75±0.02			1.94±0.03	
50-65	2.47±0.06	1.22±0.04		213±10	3.1±0.2	
60-75	2.25±0.13	2.88±0.11		126±14	1.0±0.1	

Table 4A-3: Relaxation Kinetics Probed in the Sheet of Loop 1 (1633 cm^{-1}) of FBP28 1L and FBP28 2L

FBP28 1L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	0.90±0.03	0.94±0.02	0.66±0.01	163±9	2.12±0.08	140±4
20-35	1.20±0.02	1.19±0.02	1.29±0.01	147±6	2.71±0.09	46.2±0.9
25-40	0.96±0.03	1.26±0.02	1.59±0.02	160±10	2.17±0.09	37.6±0.7
30-45	1.07±0.02	1.27±0.14	1.54±0.14	276±11	8.97±0.83	34.0±2.8
35-50	0.62±0.05		3.09±0.01	62±8		20.6±0.2
40-55	0.85±0.04		4.51±0.01	87±7		14.8±0.1
45-60	1.72±0.10		5.28±0.02	76±7		11.8±0.2
50-65	1.01±0.09		4.27±0.02	73±10		7.8±0.1
60-75	1.36±0.08		3.64±0.03	126±13		4.6±0.1
FBP28 2L						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	1.33±0.03	1.63±0.02	1.04±0.02	117±5	1.52±0.04	250±9
20-35	0.86±0.04	1.17±0.03	1.22±0.39	146±12	1.41±0.06	147±35
25-40	1.06±0.03	1.35±0.03	0.86±0.01	338±18	3.50±0.14	119±5
30-45	1.50±0.03	1.98±0.02	0.81±0.02	169±7	2.69±0.08	44.1±1.9
35-50		1.83±0.08	0.81±0.05		2.07±0.01	17.6±4.2
40-55	1.54±0.07	2.89±0.18	1.13±0.19	106±9	1.69±0.12	7.9±1.3
45-60		3.56±0.02			1.85±0.03	
50-65	2.53±0.09	3.44±0.07		154±11	2.07±0.08	
60-75	5.72±0.18	2.13±0.18		217±12	1.54±0.17	

Chapter 5: The Role of Electrostatic Interactions in Turn Stability of β-Proteins

5.1 Abstract

Electrostatic interactions play an important role in protein folding, ligand binding, and signal propagation across membranes. These charged side chains are often found in solvent-exposed regions of proteins, such as the turns of β-proteins. Here we have studied the role of charged residues in the turns of WW domains on stability and folding kinetics. The folding of wild type Pin1 WW domain, which has two positively charged residues in the first turn, was compared to mutants that incorporate a negative charge in the turn and remove the positively charged residues. A combination of FTIR spectroscopy and laser-induced temperature-jump coupled with infrared spectroscopy was used to probe changes in the amide I region. The relaxation dynamics of the peptide backbone, β-sheets and β-turns, and aspartic acid side chain of the FIP35 Pin1 mutant were measured independently by probing the corresponding bands assigned in the amide I region. Incorporation of a negative charge in the first turn was found to stabilize the peptide. The protonation state of aspartic acid is coupled to protein folding; the apparent pKa of aspartic acid in the folded protein is 6.4. Folding is initiated in the turns and the β-sheets

forming last. The dynamics of the aspartic acid follow the dynamics of the intermediate phase, supporting assignment of this to formation of the first hairpin. These results demonstrate the importance of electrostatic interactions in turn stability and formation of extended β-sheet structures**.**

5.2 Introduction

The turns of β-proteins have been shown to play a key role in modulating protein stability and folding kinetics. Hydrogen bonding, hydrophobic collapse and side-chain packing are all essential elements of β-hairpin formation.[38, 41, 47, 48, 50, 60] Charged side chains, often found in solvent-exposed regions such as the turns of proteins, can engage in ionic interactions. Electrostatic interactions play an important role in protein folding, ligand binding, and signal propagation across membranes. Salt bridges have been extensively studied for their contribution to protein stability, and have been shown to have kinetic control over the folding pathway.[125-127] Environmental pH can modulate the interactions between side chains of amino acid residues by changing their protonation states.

Here we have examined the role of electrostatics in the turn of one of the fastest folding β-protein families, the WW domain. The WW domain family consists of an antiparallel and highly twisted three-stranded β-sheet structure with a small hydrophobic core and two highly conserved tryptophan residues.[73-76] Because of their fast folding rates and simple structure, WW domains have been the focus of extensive computational and experimental studies.[70, 77-88] Many of these studies have predicted a folding mechanism where folding is initiated in the turn of the first hairpin.[70, 78, 79, 84, 86, 88, 103, 128] A comparison

Table 5-1: WW Domain Sequences and Stability*

WW Domain	Sequence	Tm (°C)
WT Pin1	KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG	59[129]
R12A WT Pin1	KLPPGWEKRMS**A**SSGRVYYFNHITNASQWERPSG	59[130]
R16A WT Pin1	KLPPGWEKRMSRSSG**A**VYYFNHITNASQWERPSG	51[130]
FIP35 Pin1	KLPPGWEKRMSR-**D**GRVYYFNHITNASQ**F**ERPSG	78[87]
R12A FIP35 Pin1	KLPPGWEKRMS**A**-**D**GRVYYFNHITNASQ**F**ERPSG	85[87]
GTT35 Pin1	KLPPGWEKRMSR-**D**GRVYYFNHIT**GTT**Q**F**ERPSG	89[77]
WT FBP28	GATAVSEWTEYKTADGKTYYYNNRTLESTWEKPQELK	64[84]
N,Q-FBP28	GATAVSEW**Q**EYKTA**N**GKTYY**Q**N**Q**RT**Q**E**Q**TWEKPQELK	55[70, 99]

*β-sheets are underlined, Mutations are bold, Turn residues with positive (blue) and negative charges (red)

of the sequence and stability of these WW domains reveals the presence of several charged residues in the first turn (Table 5-1). In the native sequences, the first turn is flanked on each side by amino acids with positively charged side chains (blue). WW domains which also have an amino acid with a negatively charged side chain (red) in the first turn have a melting temperature of ~10 °C higher than related sequences. One possibility is that these charged side chains form a salt bridge that stabilizes the first turn. This interaction is likely dependent on the protonation state of the negatively charged side chain.

We have addressed this by studying the role of aspartic acid and arginine in the Pin1 family of WW domains (Figure 5-1). WW domains have been found to be resistant to

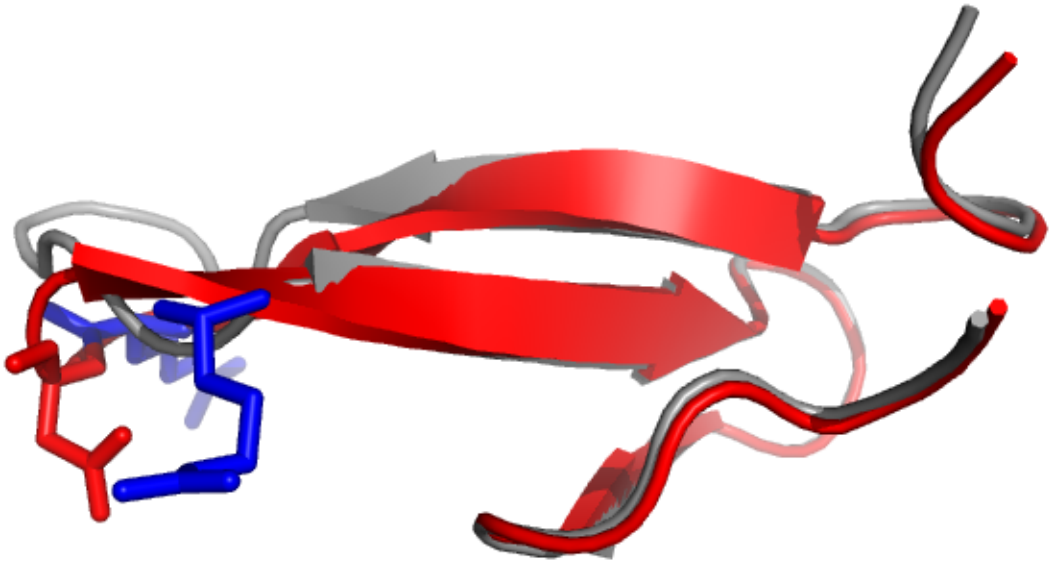


Figure 5-1. Aligned three-dimensional models of Pin1 (gray) and FIP35 (red) WW Domains obtained by x-ray diffraction. Arginines (blue) and aspartic acid (red) in the FIP35 sequence are displayed as sticks. (PDB entry: 1PIN, 2F21). The figure was prepared using PyMOL (www.pymol.org).

mutation after modification at nearly any position.[78] Mutation of the two flanking positively charged residues, arginines, has been performed with varying success.[87, 130] The arginine at the start of the turn, Arg12, has been shown to be more amenable to mutation than the arginine at the end of the turn, Arg16, likely because the side chain of Arg12 evolved for a functional role, ligand binding, rather than for protein stability. In Pin1 mutation of Arg12 to alanine or glycine resulted in no change in the stability.[130] Mutation of Arg12 to alanine in FIP35 results in a 7 °C increase in melting temperature (Table 5-

1).[87] FIP35 has a shorter turn sequence than Pin1, so it is not surprising that mutations have a slightly different effect on its stability. Mutation of Arg16 in Pin1 to alanine or glycine was slightly destabilizing, resulting in an 8 °C decrease in melting temperature.[130] Based on these observations, we conclude that mutation of the arginines to alanine is conservative and will preserve the secondary structure of the WW domain. Therefore, we investigated three WW domains: one without a negatively charged residue in the turn (Pin1), one with both negative and positively charged residues in the turn (FIP35), and one without the positively charged flanking residues in the turn (R12A R16A ΔΔFIP35).

The stability of Pin1, FIP35 and ΔΔFIP35 WW domains was determined by equilibrium circular dichroism and FTIR measurements. We found that mutation of Arg12 and Arg16 resulted in destabilization of the peptide and an increase in propensity to aggregate. Aspartic acid was protonated under folded conditions in both FIP35 and ΔΔFIP35 WW domains, making it unlikely that arginine and aspartic acid form a salt bridge. The dynamics of FIP35 WW domain formation was measured using temperature jump, time resolved infrared spectroscopy. Pulsed laser excitation was used to rapidly initiate a shift in the folding equilibrium. The relaxation of the WW domain was measured by independently probing the components of the IR amide I band assigned to the β-sheets and β-turn and the protonated carboxyl side chain of aspartic acid. FIP35 exhibited multi-exponential behavior when the carbonyl backbone was probed, but only one phase when the aspartic acid was probed.

Using infrared spectroscopy, it is possible to probe the folding dynamics of the aspartic acid with single-residue specificity without incorporating any labels. The dynamics of the aspartic acid follow the dynamics of the intermediate phase when the backbone is probed,

supporting the assignment of this phase to formation of the first hairpin. Similar to other WW domains, wavelength dependent infrared measurements in the amide I region reveal a fast 100 ns phase, intermediate ~10 μs phase and a slower ~100 μs phase.[70, 88] While the global folding mechanism agrees with computer simulations that predict folding that initiates in the first turn, these simulations methods were incapable of modeling protonation changes necessary to model the local environment of the aspartic acid in the turn. These results demonstrate the importance of electrostatic interactions in turn stability and formation of extended β-sheet structures.

5.3 Experimental Section

Protein Synthesis and Purification. FIP35, ΔΔFIP35, and Pin1 were synthesized via standard 9-fluorenylmethyloxycarbonyl (Fmoc) base solid-phase chemistry on a Liberty1 microwave peptide-synthesizer (CEM, Matthews, NC). Fmoc-PAL-PS resin (Applied Biosystems, Foster City, CA) was used to form a peptide amide. The peptide was purified by reverse-phase chromatography (C18 column) using a water/acetonitrile gradient with 0.1% trifluoroacetic acid (TFA) as the counter-ion. TFA interferes in the Amide-I IR measurements at 1672 cm^{-1}. The peptide was lyophilized and dissolved in a 2 mM HCl solution to allow exchange of the TFA counter-ion for HCl.[90] The identity of the peptide was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. The peptide was lyophilized and dissolved in D_2O to allow deuterium-hydrogen exchange of amide protons. The peptide was lyophilized a second time and resuspended in D_2O buffer. Buffers were prepared at 20 mM concentration for the appropriate pH range: acetate (pH 4-5.5), potassium phosphate (pH 6-8), Tris-HCl (pH

8.5-9), and carbonate-bicarbonate (9.5-10.5). Sample concentrations of 0.05-2.0 mM were prepared for CD, fluorescence and IR experiments.

CD Spectroscopy. CD wavelength scans and CD melting curves were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Jasco temperature controller module (Jasco, Inc., Easton, MD). Peptides were dissolved at 50 μM in buffer. All measurements were obtained using a 1-mm pathlength cell. Wavelength scans were recorded over the range of 260 to 190 nm with an average of 3 repeats. During spectral acquisition, a bandwidth of 2 nm and scan rate of 50 nm/min was used. Thermal unfolding experiments were performed by monitoring the signal at 226 nm from 5 to 95 °C using a 0.1 °C interval and scan rate of 30 °C/hr. During the thermal unfolding experiment a full wavelength scan was obtained every 5 °C after a 60 second delay. The buffer and protein concentrations were the same as used in the wavelength scan experiments.

FTIR Spectroscopy. The equilibrium melting behavior was monitored on a Varian Excalibur 3100 FTIR spectrometer (Varian Inc., Palo Alto, USA) using a temperature controlled IR cell. The IR cell consists of two CaF_2 windows stacked and separated by a 100 μm Teflon spacer split into two compartments, a sample and a reference. The same cells are used for equilibrium FTIR and T-jump experiments. No aggregation was observed in the infrared at reported concentrations. All spectra shown at a specific temperature are constructed by subtracting the spectrum of reference buffer solution without protein from sample solution with protein. The temperature-dependent difference spectra were then generated by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures. Data collected at 50 μM concentration were baseline

corrected using a baseline spline fit. The second derivative spectra were computed in IGOR PRO after smoothing the data with a sixth order binomial algorithm to remove any residual water vapor (WaveMetrics, Lake Oswego, OR).

Singular Value Decomposition and Global Fitting. Singular value decomposition (SVD) was used to determine the number of spectral components required to describe the temperature dependence of the protein FTIR and CD spectra.[131] The SVD and global fit methods have been described previously.[132] The temperature dependent data was used to construct a data matrix, A, where each column represents the spectra of interest at a specific temperature. SVD analysis of the data results in three matrices, $A=USV^T$, where U, S, and V^T are matrices containing the basis spectra, singular values and temperature evolution of the basis spectra, respectively. Basis spectra containing only noise were identified by their singular values and eliminated from the set of basis spectra. Multiple basis spectra were necessary to reconstruct the original data matrix, A.

A global fit was used to separate the individual spectral components in the basis spectra and their temperature dependence. The matrix A can be described by $A=USV^T$ or $A=DF^T$, where D and F^T are matrices containing the spectral components and temperature dependences, respectively. Therefore, $D=USV^TF^{T+}$ where F^{T+} is the pseudoinverse of F^T. This can be rewritten D=USH, where $H=V^TF^{T+}$. The H matrix can be determined through a global fit of the V^T matrix using a mathematical model that consists of one sigmoidal transition for each spectral component:

$$f(T)_1 = h_{11} * \left(b_1 + \frac{m_1}{1+exp\left(\frac{(T_{m_1}-T)}{\Delta_1}\right)} \right) + h_{12} * \left(b_2 + \frac{m_2}{1+exp\left(\frac{(T_{m_2}-T)}{\Delta_2}\right)} \right) + \cdots$$
$$f(T)_2 = h_{21} * \left(b_1 + \frac{m_1}{1+exp\left(\frac{(T_{m_1}-T)}{\Delta_1}\right)} \right) + h_{22} * \left(b_2 + \frac{m_2}{1+exp\left(\frac{(T_{m_2}-T)}{\Delta_2}\right)} \right) + \cdots \quad (5\text{-}1),$$
$$\vdots$$

where h_{11}, h_{12}, …, h_{1n}, h_{21},h_{22}, …, h_{2n} are the elements of an nxn H matrix and T_m and Δ represent the midpoint and width of each of the transitions. The SVD analysis and global fitting were performed in IGOR PRO.

Time Resolved Temperature Jump (T-jump) Relaxation Measurements. The IR T-jump apparatus has been described previously.[30] Pulsed laser excitation is used to rapidly perturb the folding equilibrium on a timescale faster than the molecular dynamics of interest. Time resolved infrared is then used to probe the reaction. A Q-switched GCR-4 Nd:YAG laser (Spectra Physics, Mountainview, CA) fundamental at 1064 nm is Raman shifted (one stokes shift in 200 psi H_2 gas) to produce a 10 ns pulse at 2 μm. The magnitude of the T-jump is calculated using the change in reference absorbance with temperature. The T-jump reference is taken from D_2O buffer with 20 mM potassium phosphate buffer at pD* 7.0 at the same temperature and frequency as the sample. Absorbance changes at the reference frequency are due only to changes in D_2O absorbance, which is used as an internal thermometer.[30]

The change in signal induced by the T-jump is probed in real time by a continuous laser with a frequency in the amide I' band of the IR. The mid-IR probe beam is generated by a continuous wave quantum cascade laser (Daylight Solutions Inc., San Diego, CA) with a tunable output range of 1570-1730 cm^{-1}. The transient transmission of the probe beam through the sample is measured using a fast, 100 MHz, photovoltaic

MCT IR detector/preamplifier (Kolmar Technologies, Newburyport, MA). Transient signals are digitized and signal averaged (1000 shots) using a Tektronics digitizer (7612D, Beaverton, OR). Instrument control and data collection are controlled using a LabVIEW computer program.

Analysis of Kinetics Data. The peptide relaxation kinetics must be deconvolved from the observed kinetics. Accurate deconvolution is possible as the instrument response is determined from the reference measurement under the exact conditions of the sample measurements. In order to minimize detector artifacts, the reference is scaled prior to subtraction from the sample. The decay function is a multi-exponential decay with the formula:

$$A = A_0 + \cdots + A_n exp\left(\frac{-(x-x_0)}{\tau_n}\right) \quad (5\text{-}2),$$

where A_0 is an offset, n is the number of exponentials to fit, A_n is a preexponential factor, τ_n is the relaxation lifetime of the sample and x_0 is the time offset. In order to best fit the data, the minimum number of exponentials with unique relaxation lifetimes was selected. The data analysis was performed in IGOR PRO (WaveMetrics, Lake Oswego, OR).

Results and Discussion

5.4 Far-UV CD Spectroscopy. The Pin1, FIP35 and ΔΔFIP35 WW domains were synthesized to study the aspartic acid in loop 1 of FIP35 and its interaction with neighboring arginine residues. Far-UV CD is a good indicator of peptide secondary structure. Typically, β-sheet peptides have a CD spectrum with a negative peak at ~218 nm and a positive peak at ~195 nm.[91] Folded WW domains instead have a CD spectrum dominated by a negative peak at ~202 nm and a positive peak at ~230 nm.[92, 93] Small

variations in the position and intensities of these peaks have been observed amongst the WW domain family.[93] The peak at ~202 nm resembles the random coil peak usually found at 200 nm. Disorder in the N- and C- termini of the folded structure of WW domains is thought to contribute to the negative peak in the CD spectrum.[92, 94-96] The peak at ~230 nm arises from the presence of ordered aromatic side chains.[79, 91, 92] Folded Pin1 WW domains and its mutants have been shown to have additional structure between 190-210 nm.[96, 130, 133] The far-UV CD spectra of the Pin1, FIP35 and ΔΔFIP35 WW domains are consistent with the Pin1 family WW domain secondary structure (Figure 5-2A).

Thermal denaturation was monitored by recording the ellipticity change at 226 nm with temperature (Figure 5-2B). The Pin1 and FIP35 WW domains exhibit the typical heat induced unfolding behavior with a loss of intensity at 226 nm and a shift of the minimum at 197 nm to 200 nm, corresponding to a change in the secondary structure from WW domain to random coil (Figure 5-2A).[92] In addition to the peaks associated with thermal unfolding, the ΔΔFIP35 mutant gains a negative peak at ~219 nm and positive peak at ~190 nm, consistent with β-sheet structure. This is evidence for formation of β-sheet aggregates at high temperature. The melting curves (Figure 5-2B) were fit to an apparent two-state equilibrium model:

$$A_0 = \frac{A_f}{1+exp\left(-\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)\right)} + \frac{A_u}{1+exp\left(\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)\right)} \qquad (5\text{-}3),$$

where A_0 is the observed absorbance, A_f and A_u are the absorbance contributions from the folded and unfolded populations, ΔH is the enthalpy change at the midpoint, R is the gas constant, and T_m is the transition midpoint.[134] This analysis assumes a ΔC_p of 0 as it is

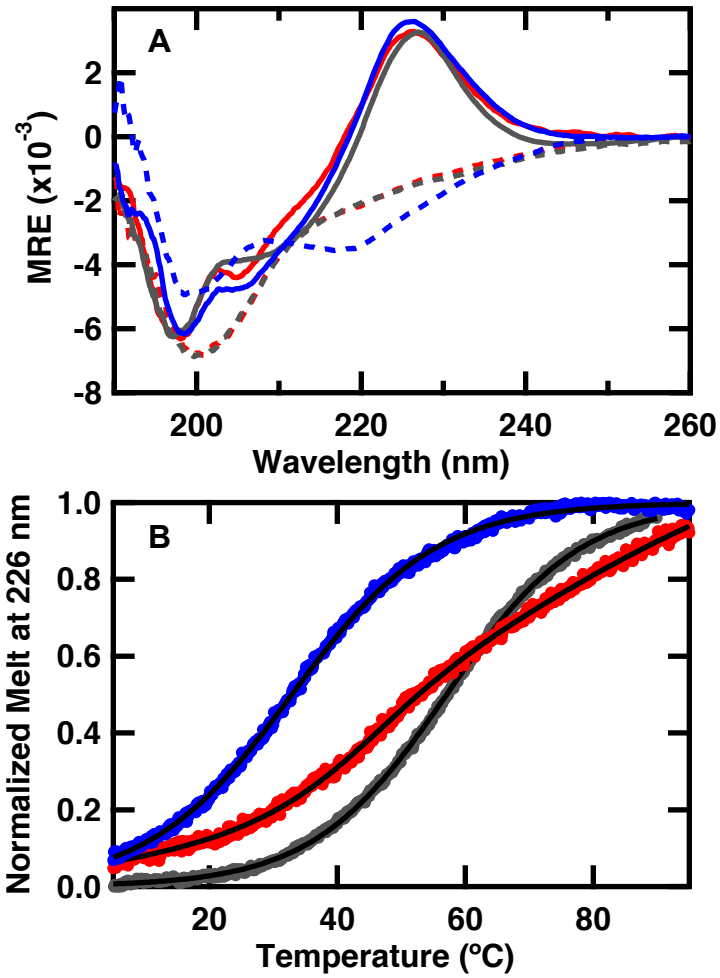


Figure 5-2. (A) Far-UV CD spectra of 50 μM solutions of the FIP35 (red), ΔΔFIP35 (blue) and Pin1 (gray) WW domains in 20 mM potassium phosphate buffer (pH 7.0) acquired at 5 °C (____) and 95 °C (----) during the course of a thermal denaturation in a 0.1 cm pathlength cell. (B) Thermal denaturation of the WW domains monitored by CD at 226 nm. The continuous line represents the best fit to the data to a sigmoid (Equation 5-3).

unlikely that a small peptide like a WW domain would have a large difference in heat capacity between the folded and unfolded states. The data are then normalized for comparison. The observed melting temperature of Pin1 is 57.1 ± 0.1 °C. This agrees with the previously reported melting temperature of Pin1, 58.6 °C, obtained by a CD melt monitored at 226 nm.[133] An SVD analysis of the FIP35 FTIR spectra collected during the thermal melt reveals that there are two dominant subspaces. The FIP35 melting curve was fit to the sum of two melts with a melting temperature of 71 ± 3 °C, the dominant subspace, and 45.1 ± 0.4 °C. The previously reported melting temperature of FIP35, 78 °C, was obtained by a CD melt monitored at 227 nm.[87] The slight difference in the T_m is likely due to error in fitting the two subspaces simultaneously. The melting temperature of ΔΔFIP35 is 32.9 ± 0.1 °C. Mutation of the two arginines in loop 1 of FIP35 to alanine results in a 40 °C destabilization. The breadth of the transition is similar to Pin1 demonstrating that while mutation of loop 1 destabilizes the protein, the cooperativity of the folding transition is unchanged (Figure 5-2B).

5.5 FTIR Spectroscopy. The temperature-induced unfolding of Pin1 and FIP35 was studied over the range from 10 to 90 °C in 5 °C intervals using FTIR spectroscopy monitored in the amide I' region. To avoid aggregation of ΔΔFIP35 observed at high temperature by CD, the FTIR absorption spectrum was collected at 50 μM at 20 °C. The high and low temperature absorption spectra of the amide I' spectral region (amide I region of peptides in D_2O) of the peptides are shown in Figure 5-3A (complete temperature dependent spectra are shown in Supporting Information Figure 5A-1). The

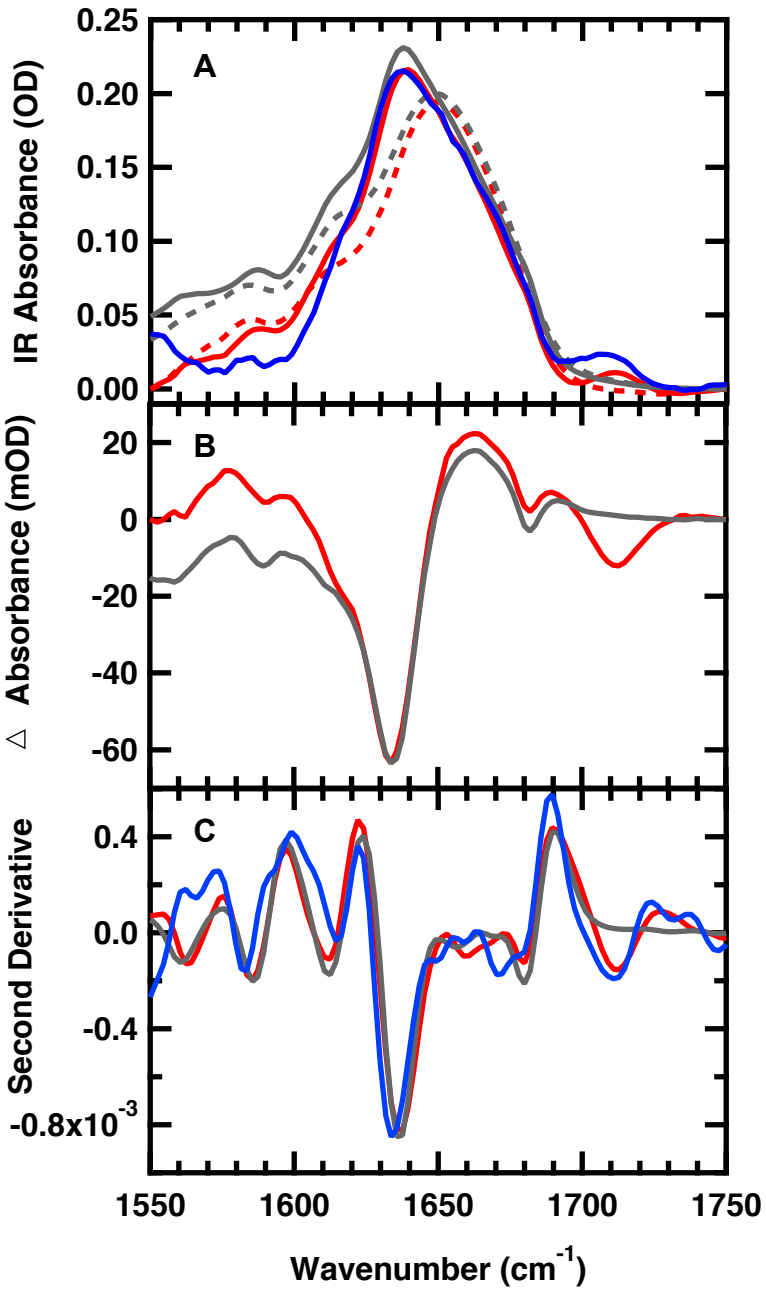


Figure 5-3. Temperature dependent FTIR spectra of 1.5 mM FIP35 (red), 50 μM ΔΔFIP35 (blue) and 1.5 mM Pin1 (gray) WW Domains in 20 mM potassium phosphate buffer (pH 7). (A) Absorbance spectra in the Amide I' region; acquired at 20 °C (____) and 75 °C (----) during the course of a thermal denaturation. ΔΔFIP35 data is normalized

at 1636 cm^{-1} for comparison. (B) Difference spectra obtained by subtracting the spectrum at 20 °C from the spectra at 75 °C. Data is normalized at 1636 cm^{-1} for comparison. (C) Normalized second derivative of FTIR spectra at 5 °C.

amide I' absorbance arises from C=O stretching vibration of the polypeptide backbone carbonyls, and is an established indicator of secondary structure.[62-64] This relatively broad band contains contributions from the entire polypeptide backbone, which in the case of WW domains includes β-sheet, β-turn and random coil structure. The changes with temperature are highlighted by the difference spectra for each peptide (Figure 5-3B). The difference spectra are generated by subtracting the lowest temperature spectrum from each absorption spectrum at higher temperature. Negative peaks correspond to specific structures or interactions present in the folded state, and positive peaks correspond to new interactions with solvent in the unfolded state. The individual peaks are more easily distinguished in the second derivative of the FTIR spectra at the lowest temperature (Figure 5-3C).

At low temperature there are three main components of the amide I' band, centered at 1613, 1636 and 1680 cm^{-1}. The intensity of these features decreases with increasing temperature, meaning that they are all associated with the folded state. These peaks have previously been observed in other WW domains.[70, 88, 100] A peak at ~1611 cm^{-1} in β-hairpins has been assigned to an amide C=O group in the turn usually involved in multiple hydrogen bonds with side chain or backbone donors.[50, 66] There are three such groups in the first turn and one such group in the second turn of wildtype Pin1 WW domain.[135] Upon mutation of the first turn to SA-DGR, a FIP35 like mutant, one of the

interactions in the first turn and the interaction in the second turn are conserved, and two new hydrogen bonds between backbone carbonyls and side chain or backbone donors arise.[117] The peak is of equal intensity in each of the peptides (Figure 5-3C), because the number of hydrogen bonds is conserved. IR bands at 1634 and 1631 cm^{-1} are well established components of antiparallel β-sheets.[102] The peak at 1634 cm^{-1} arises from in-phase coupling of carbonyl groups in the sheet and the peak at 1681 cm^{-1} arises from out-of-phase coupling of carbonyls in the sheet. The Pin1, FIP35 and ΔΔFIP35 amide I' bands at 1636 and 1680 cm^{-1} are consistent with these characteristic β-sheet markers.

Analysis of the spectral region above 1700 cm^{-1} reveals a peak at 1713 cm^{-1} that is present in the FIP35 and ΔΔFIP35 WW domain, but not the Pin1 WW domain. This peak arises from the C=O stretch of a protonated carboxyl side chain, either glutamic acid or aspartic acid.[40] The difference between the residues with carboxyl group side chains in the three sequences is in the first turn; FIP35 and ΔΔFIP35 have an aspartic acid that is not found in Pin1. Therefore, the peak at 1713 cm^{-1} can be assigned to aspartic acid 13. A deprotonated carboxylate group has two strong bands at 1400 and 1570 cm^{-1} (ν_{sym} and ν_{asym}). These positions may shift upon cation chelation, which changes bond lengths and bond angles. In unidentate chelation of the deprotonated carboxylate group, the two CO bonds are no longer equivalent, with the chelating oxygen having more single bond character and the non-chelating oxygen having more double bond character. In extreme cases this has been found to shift one of the bands to 1713 cm^{-1}, the band assigned to the C=O stretch of a protonated carboxyl side chain.[136] There are two arginines in FIP35 that may provide a cation through a salt bridge with aspartic acid, arginine 12 and arginine 15, leading to this phenomenon. Mutation of these arginines to alanine, ΔΔFIP35, does not

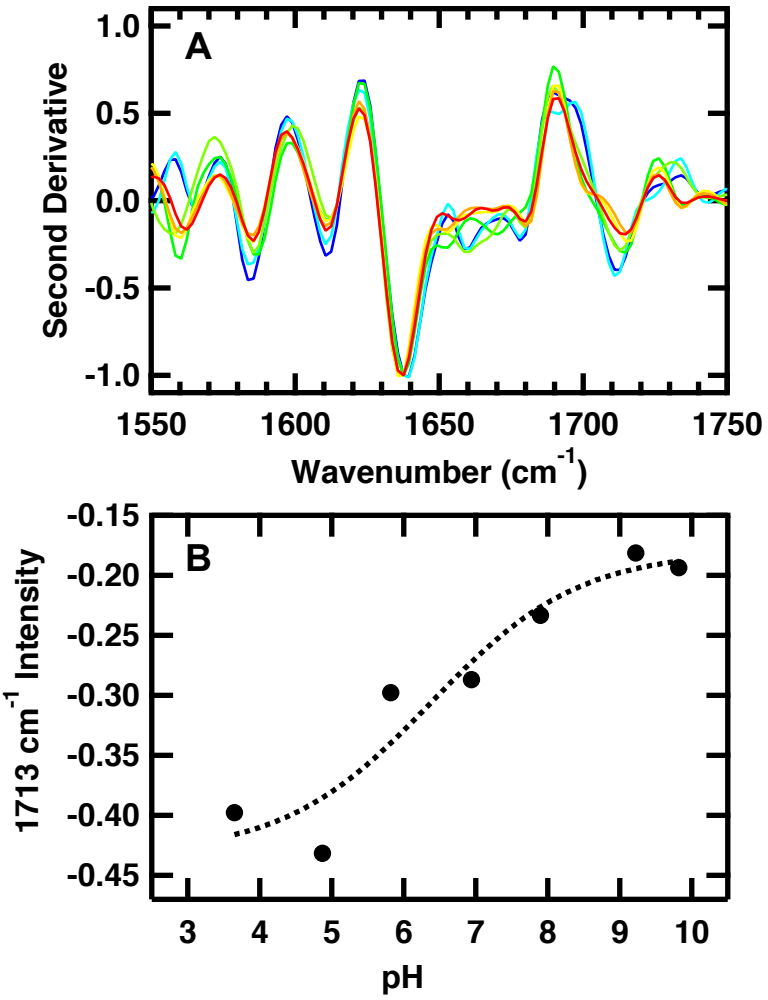


Figure 5-4. pH dependent FTIR spectra of 0.8 mM FIP35 WW Domain in buffer (see experimental section for buffer pH 4-7) at 20 °C. (A) Normalized second derivative of FTIR spectra pH 4-7 (blue-red). (B) pH titration of protonated carboxylic acid obtained by plotting the change in IR second derivative spectra at 1713 cm^{-1} versus pH. The data are fit to an apparent 2-state model (Equation 5-3).

eliminate the peak at 1713 cm^{-1} (Figure 5-3), supporting our assignment of this band to the protonated carboxylic acid side chain of aspartic acid 13.

The pH dependence of the protonated carboxylic acid side chain was studied over the range from pH 4-10 (Figure 5-4). The pH was not dropped below pH 4 to minimize contributions to the 1713 cm^{-1} band from the protonated glutamic acid side chain (pK_a of 4.2), which overlaps with the protonated aspartic acid side chain band. Measurements were carried out at 20 °C, a temperature where at neutral pH the peptide was found to be folded (Figure 5-3C). The three bands assigned to the folded WW domain, 1613, 1636 and 1680 cm^{-1}, are observed at each pH, confirming that FIP35 is folded over the entire pH range (Figure 5-4A). The aspartic acid is protonated to some extent over the entire pH range. An increase in the intensity of the turn structure, 1613 cm^{-1}, correlates with the increase in the population of protonated aspartic acid, 1713 cm^{-1}, as the pH is lowered. This is evidence that changes in the protonation state of aspartic acid, found in the first turn, affect the hydrogen-bonding network of the first turn. A pH titration of the protonated carboxylic acid side chain of aspartic acid derived from the second derivative of the pH dependent IR absorbance at 1713 cm^{-1} is shown in Figure 5-4B. The pK_a derived from this fit is 6.4 ± 0.9. The pK_a of the aspartic acid side chain of free aspartic acid is 3.9; however, electrostatic and hydrophobic induced pK_a shifts of as many as 6 pK_a units have been reported for aspartic acid in proteins.[137-140] The transition is broad, likely because protonation is coupled with folding of the WW domain. At neutral pH the

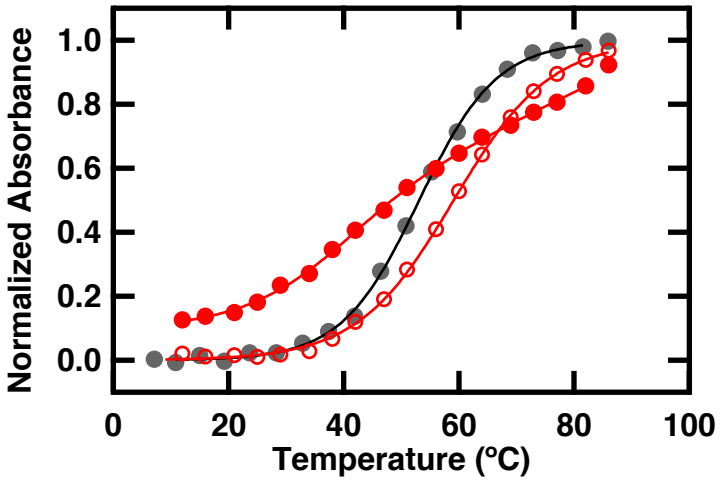


Figure 5-5. FTIR melt curves for the FIP35 (red) and Pin1 (gray) WW domains obtained by plotting the change in IR difference spectra at 1636 cm^{-1} (closed circles) and 1713 cm^{-1} (open circles) versus temperature. The data are fit to an apparent 2-state model and then normalized (Equation 5-3).

aspartic acid completely deprotonates when the WW domain unfolds (Fig 5-3A). The pK_a of the aspartic acid 13 side chain is dependent on both pH and the FIP35 WW domain secondary structure.

The normalized melting curves for Pin1 and FIP35 WW domains derived from the temperature dependent IR absorbance at 1636 cm^{-1} and 1713 cm^{-1} are shown in Figure 5-5. The data were normalized after being fit to an apparent two-state equilibrium model using Equation 5-3. The melting temperature of the wildtype Pin1 is 53.1 ± 0.2 °C. The melting temperature of FIP35 is wavelength dependent, with a relatively broad transition when probed at 1636 cm^{-1} and a sharper transition when probed at 1713 cm^{-1}. When the

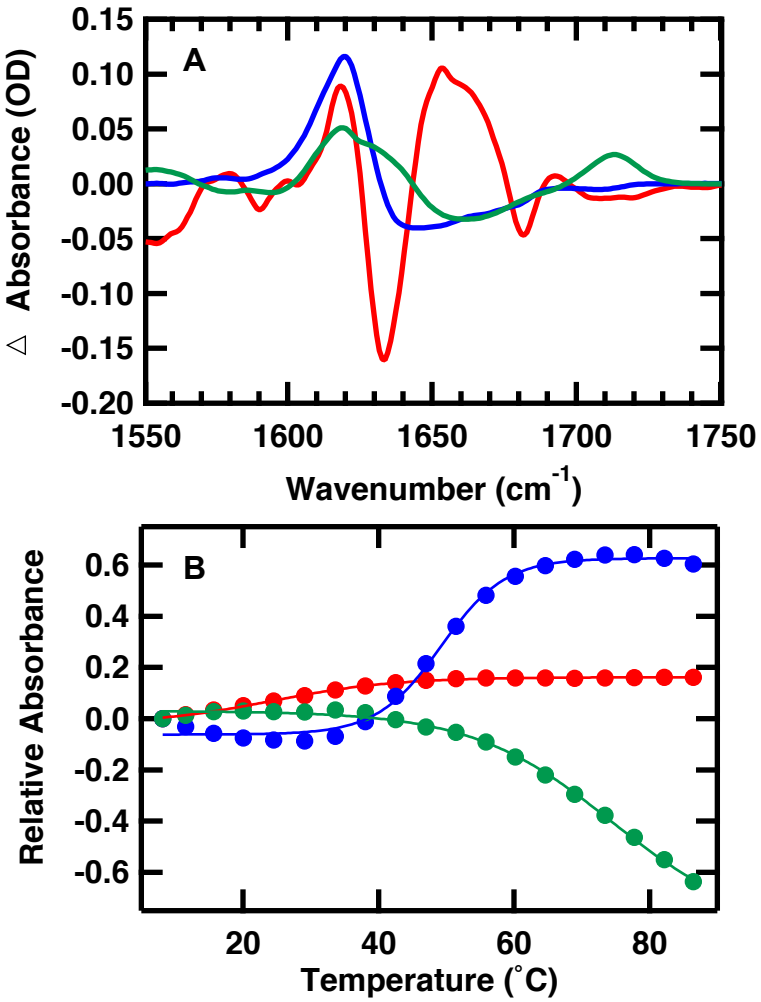


Figure 5-6. The D-spectral components (A) and the corresponding F^T-temperature profiles (B) determined from a SVD and global fitting analysis of the temperature-dependent FIP35 FTIR difference spectra. Component 1 is red, component 2 is blue and component 3 is green.

1636 cm^{-1} data is fit to the sum of two melts it fits to a T_m of 40.4 ± 0.1 °C and 77.1 ± 0.1 °C. This agrees well with the melting temperature probed by CD. The transition at

1713 cm^{-1} fits to a T_m of 59.1 ± 0.2 °C. The observation of frequency dependent melt profiles suggest the presence of more than one temperature-dependent process. The single frequency melt curves are less accurate than curves obtained from SVD analysis, because the latter are obtained using the full frequency dependence of the absorption spectra.

A combined SVD and global fitting protocol applied to the temperature dependent FIP35 spectra shows three statistically significant spectral components. The three D-matrix spectral components are show in Figure 5-6A with the corresponding temperature dependences (F^T-matrix) shown in Figure 5-6B. The F^T-temperature profiles for the three D-spectral components exhibit an apparent Tm of 24.4 ± 0.8, 48.7 ± 0.2 and 74 ± 2 °C. The first spectral component resembles the difference spectra in Figure 5-3B, with negative bands at 1636 and 1680 cm^{-1} arising from the loss of β-sheets and positive band at 1660 cm^{-1} corresponding to new interaction with solvent at higher temperature. There is an additional peak at 1619 cm^{-1}, which could indicate conversion of 3-stranded β-sheet to 2-stranded β-sheet or formation of parallel aggregates. The corresponding F^T temperature profile shows that the transition is broad, likely representing melting of the sheets prior to the global transition at 74 °C. The second D-spectral component contains a dominant band at 1620 cm^{-1} due to formation of parallel β-aggregates.[141] It is possible to eliminate FIP35 aggregation by collecting IR data at fewer temperatures (Figure 5-3), thereby reducing the time the sample is at high temperature, or by lowering the pH (Figure 5A-2). The corresponding F^T-temperature profile shows that the number of parallel aggregates increase as the temperature is raised above 48.7 °C. No aggregation was observed during kinetic measurements, even with T-jumps to temperatures higher than 49°C. This suggests that the aggregation time scale is significantly longer than the

duration of the T-jump, which relaxes back to the initial temperature in a few milliseconds. The third D-spectral component has three peaks, centered at 1619, 1629 and 1713 cm^{-1}. Peaks at 1619 and 1629 can be assigned to the turn and β-sheet of a single hairpin, respectively.[41] The 1713 cm^{-1} peak can be assigned to the aspartic acid found in the first turn. From the corresponding F^T temperature component we conclude that this is the melt of the first loop of FIP35, the more stable hairpin of the WW domain. From combined SVD and global analysis we observe melting of the third sheet over a broad range of temperatures whereas the global transition occurs with the melt of the first loop.

5.6 Temperature-Jump Relaxation Kinetics. The relaxation kinetics of the folding/unfolding transition of FIP35 WW domain following a laser induced temperature jump were probed using time-resolved infrared spectroscopy. Structure specific measurements were made using the amide I' frequency for the amide bond vibration of the turn (1613 cm^{-1}), residues involved in coupling between the strands of the sheets (1636 and 1680 cm^{-1}), and the protonated carboxylic acid side chain of aspartic acid (1713 cm^{-1}). Jumps were performed slightly off peak center to maximize the transient absorbance signal. The complete relaxation kinetics for each frequency is reported in the supporting information (Table 5A-1, 5A-2, 5A-3, 5A-4). Time resolved measurements examined the dependence of the relaxation rates on the final temperature following a temperature jump. The magnitude of the temperature jump was kept constant while varying the final temperature for a range of final temperatures below the global melting transition. The relaxation kinetics of the amide I' region (1613, 1636 and 1680 cm^{-1}) are best fit by a triple exponential (Equation 5-2). At high temperatures the 1680 cm^{-1} data is

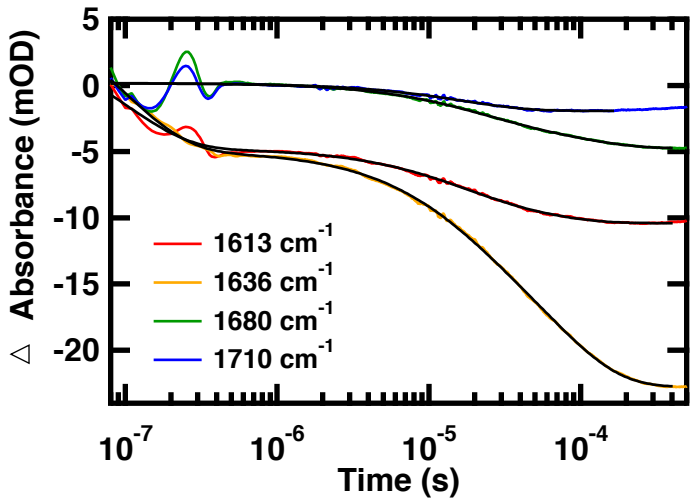


Figure 5-7. Representative IR T-jump relaxation kinetics of FIP35 monitored in the amide I' spectral region at 1613, 1636, and 1680 cm^{-1} and the protonated carboxylic acid side chain of aspartic acid at 1710 cm^{-1} following a T-jump from 35 to 50 °C. A triple exponential fit is overlaid on the 1613 and 1636 cm^{-1} kinetic trace, double exponential fit is overlaid on the 1680 cm^{-1} kinetic trace and a single exponential fit is overlaid on the 1710 cm^{-1} kinetic trace (Equation 5-2).

fit to a double exponential due to artifacts in the data at early time and the overall low absorbance change. The relaxation kinetics of the aspartic acid side chain is best fit by a single exponential at all temperatures. Although aggregation was observed over the timescale of equilibrium FTIR measurements at pH 7, it does not contribute to the T-jump dynamics because aggregation is slower than the duration of the T-jump, which relaxes back to the initial temperature in a few milliseconds. IR T-jump measurements at pH 5.8, where aggregation in FTIR measurements is not observed, agree with measurements obtained at pH 7 (Table 1S-4S). Figure 5-7 displays the relaxation kinetics

Table 5-2: Relaxation Kinetics Following a Jump from 35 to 50 °C

FIP35 WW Domain						
	A_1 (mOD)	τ_1 (ns)	A_2 (mOD)	τ_2 (μs)	A_3 (mOD)	τ_3 (μs)
1613 cm^{-1}	3.4 ± 0.1	108 ± 4	3.6 ± 0.2	14.7 ± 0.7	2.1 ±0.2	56 ± 6
1636 cm^{-1}	5.0 ± 0.1	95 ± 2	4.8 ± 0.1	13.9 ± 0.2	13.0 ± 0.1	70 ± 1
1680 cm^{-1}			2.5 ± 0.2	18 ± 1	2.5 ± 0.2	86 ± 8
1710 cm^{-1}			1.9 ± 0.1	17 ± 1		

of FIP35 following a jump from 35 to 50 °C. The fits of the data in Figure 5-7 are reported in Table 5-2. There is good agreement between the dynamics at each of the amide I' frequencies. There is a fast ~100 ns phase (τ_1), a ~10 μs phase (τ_2) and a slower ~100 μs phase (τ_3). The single phase observed when the frequency assigned to the aspartic acid side chain is probed agrees with the intermediate τ_2 phase.

Frequency dependent measurements of the amide I' region reveal differences in the dynamics of the turns (1613 cm^{-1}) and β-sheets (1636 and 1680 cm^{-1}). Because of overlap of peaks in the amide I' region, probes at each location are sensitive to dynamics associated with folding at multiple locations, resulting in mutiexponential kinetics regardless of probe frequency. It is also likely that the spectral responses are coupled such that formation of one structure affects the others. However, the relative amplitude of each kinetics phase depends on the probe frequency, which allows us to assign the structural feature that contributes most strongly to each phase. The relative amplitude of

the fast phase is greatest in the transient measured at 1613 cm^{-1}, which probes the turns of the WW domain. The slow phase dominates the kinetics measured at 1636 cm^{-1}, which probes interstrand coupling across the β-sheets of the WW domain. This assignment is supported by the kinetics measured at 1680 cm^{-1}, also dominated by the slow phase, which probes out-of-phase coupling of the carbonyl bonds in the β-sheet. Based on these observations we assign the fast phase to turn formation and the slowest phase to sheet formation.

Unlike the measurements in the amide I' region, the probe at 1713 cm^{-1} is sensitive to dynamics of a single side chain, aspartic acid. The peak is in a clear spectral region, easing interpretation of observed dynamics. The observed dynamics report on the folding of the first hairpin of FIP35, because the aspartic acid, located in the first turn, is only protonated when FIP35 is folded. It is likely that many of the contacts in the first hairpin must be formed in order for the environment to be suitable for the aspartic acid to protonate. The single exponential kinetics match the kinetics of the intermediate phase probed in the amide I' region (Table 5-2). An Arrhenius plot of the observed kinetics at 1713 cm^{-1} and 1636 cm^{-1} demonstrates this agreement across all of the temperatures probed (Figure 5-8). Based on this, we can assign the intermediate phase to formation of first hairpin. Using observations from the frequency dependent measurements in the amide I' region and of the aspartic acid side chain it is possible to interpret the three phases observed in the Arrhenius analysis. The fast relatively temperature independent 100 ns phase is assigned to formation of the turn. The intermediate ~10 μs phase corresponds to formation of the first hairpin and the slow highly temperature dependent ~100 μs phase can be attributed to formation of the second hairpin.

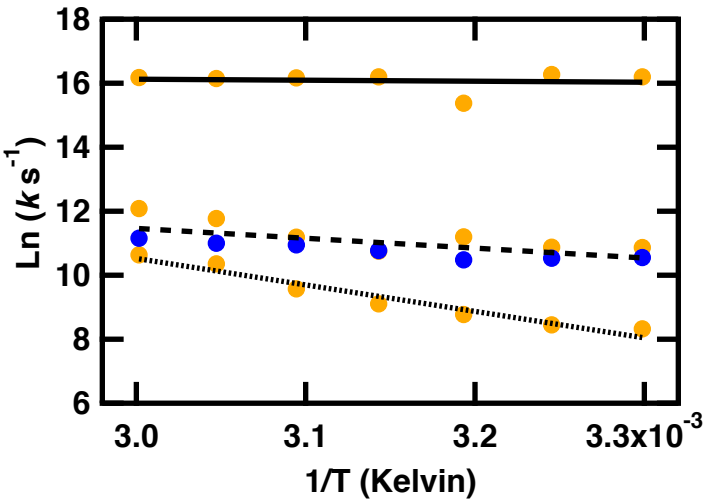


Figure 5-8. Arrhenius plot showing the temperature dependence of the folding region at 1636 cm^{-1} (orange) and 1710 cm^{-1} (blue). The values of T used for the (1/T) axis are the final temperatures reached during the jump. *k* is the value obtained from a fit of the T-jump transient of FIP35. Lines are a result of fitting τ_1 (solid line), τ_2 (dashed line), and τ_3 (dotted line) of FIP35.

Previous temperature jump studies of FIP35 probed by fluorescence spectroscopy reported biphasic behavior below T_m and single exponential behavior above Tm. A jump to a final temperature of 64 °C, below T_m, fit to relaxation lifetimes of 1.5 and 15 μs.[87] The difference in folding behavior above and below T_m was interpreted as the onset of downhill folding. The infrared measurements that we performed in the amide I' region were consistent with these phases, but with an additional ~100 ns phase. Fluorescence measurements are sensitive to the tryptophan side chain packing in the β-sheet of the first hairpin, whereas infrared measurements are sensitive to secondary structure changes in

the peptide backbone. Because the ~100 ns phase is located in the turn and the tryptophan is located in the β-sheet, the tryptophan may not report on dynamics in the turns. This phenomenon was also observed in temperature jump experiments probed by fluorescence and infrared spectroscopy in FBP28 WW domain.[70] Our results do not support a model of downhill folding, but rather a hierarchical model of folding initiated in the turns, followed by the hairpin 1 and hairpin 2. The fast phases are temperature independent and the slowest phase is highly temperature dependent, so as the temperature is raised the intermediate and slow phase overlap making it impossible to separate the two phases. Measurements of the aspartic acid side chain support this model; if multiple folding pathways existed we would expect multi-exponential kinetics when this is probed, however, the dynamics are single exponential consistent with a hierarchal model.

Extensive simulations of FIP35 predict folding is more likely to proceed through formation of the first hairpin followed by the second hairpin, however, disagree on the presence of intermediates or downhill folding.[21, 128, 142, 143] In the case of two 100 μs trajectories of FIP35 WW domain made available by D. E. Shaw Research, downhill folding or folding through intermediates is predicted depending on the analysis method.[128, 142-144] Our results support a folding model initiated in the first hairpin, which proceeds through a series of intermediates. While the global folding mechanism is in good agreement with simulation predictions, we observe changes in the protonation state of aspartic acid during folding that have not been captured by simulation methods. Optimization of the stability and folding kinetics of Pin1 WW domain was achieved with three charged residues located in the first turn. WW domains fold through an intermediate where the first hairpin forms first, so experimental and computational methods that

capture electrostatic interactions may provide a clearer understanding of the folding pathway.

5.7 Conclusion

Using infrared spectroscopy we observe a protonated side chain in the first turn of WW domains that increases the overall protein stability. Time resolved measurements probed in the amide I region and at the aspartic acid side chain provide evidence for a hierarchical folding mechanism, similar to that proposed for other WW domains.[70, 78, 86, 88, 103] Formation of the turns precedes formation of the first hairpin, with the final step being folding of the second hairpin. Infrared spectroscopy of the protonated side chain exhibits a single exponential that agrees with the intermediate phase, supporting the hierarchical folding model initiated in the first hairpin. Using infrared spectroscopy it is possible to monitor the side chain dynamics with single-residue specificity without incorporation of infrared labels. These experiments capture important changes in side chain protonation in the first turn of WW domain mutants where the first turn has been optimized for stability and kinetics.

5.8 Appendix

1. Temperature dependent FTIR spectra of Pin1 and FIP35 WW domains were recorded to confirm that the peptide was folded. The FTIR spectroscopy methodology is given in the experimental section of the full paper. Temperature dependent FTIR of Pin1 and FIP35 WW domains collected at pH 7 are reported in Figure 5A-1. Temperature dependent FTIR of FIP35 collected at pH 5.8 is reported in Figure 5A-2.

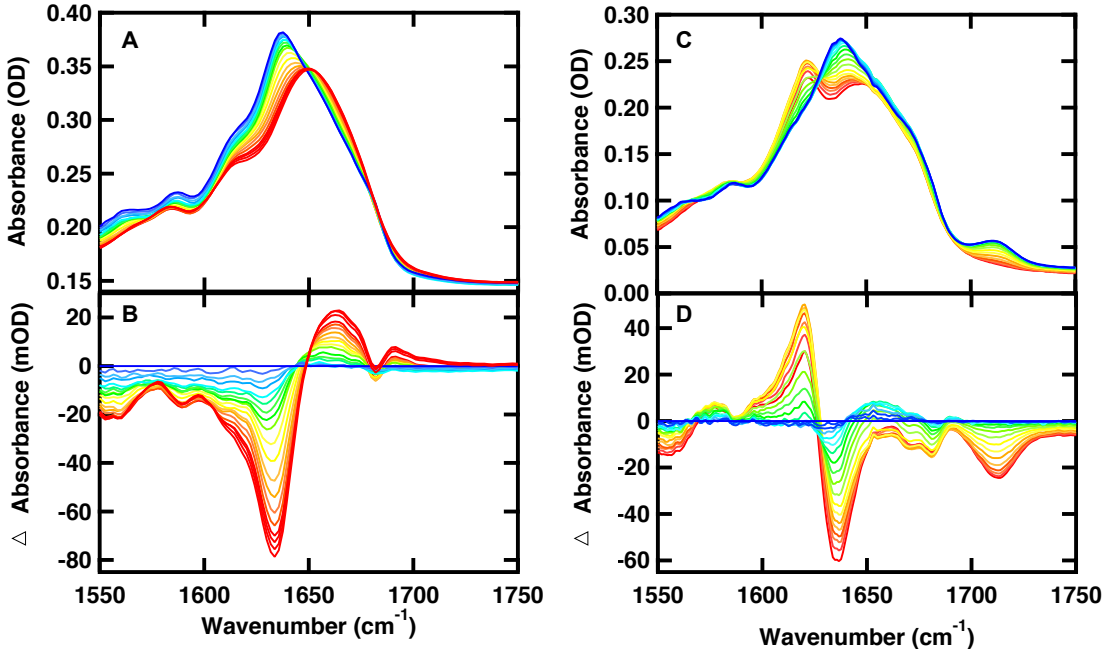


Figure 5A-1. Temperature dependent FTIR spectra of 1.5 mM Pin1 (A, B) and Fip35 (C, D) in 20 mM potassium phosphate buffer (pH 7). (A, C) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 5 to 85 °C in ~5 °C intervals. (B, D) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher temperatures.

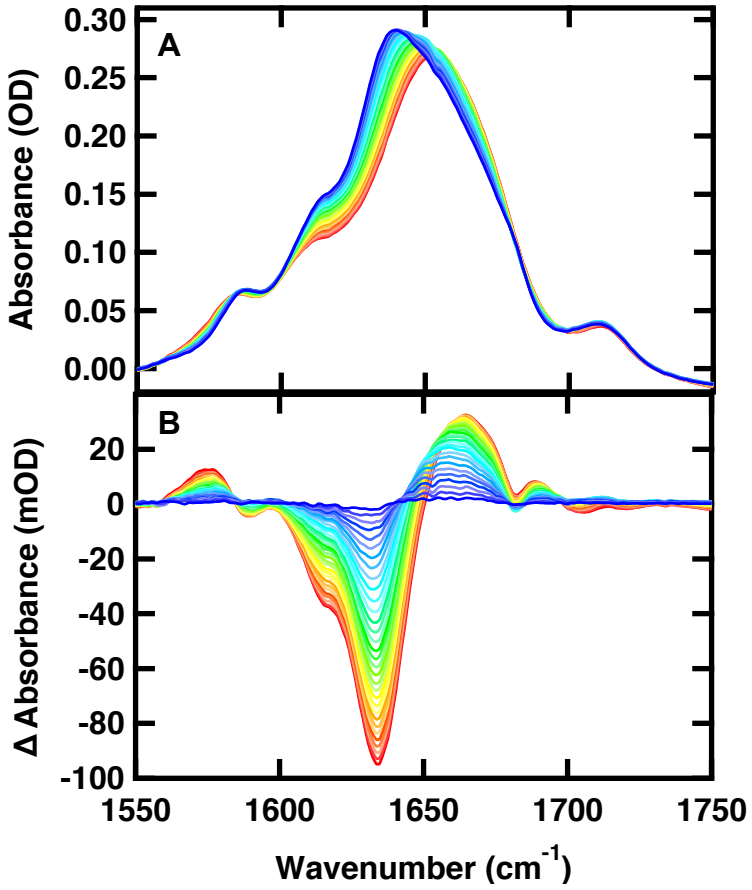


Figure 5A-2. Temperature dependent FTIR spectra of 2 mM Fip35 (A, B) in 20 mM potassium phosphate buffer (pH 5.8). (A) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 5 to 85 °C in ~5 °C intervals. (B) Difference spectra obtained by subtracting the spectrum at 5 °C from the spectra at higher temperatures.

2. Temperature dependent infrared T-jump measurements were collected to determine the dynamics of the systems. The IR T-jump methodology is given in the methods section of the full paper. Wavelength dependent IR T-jump measurements were collected at 1613 (Table 5A-1), 1636 cm^{-1} (Table 5A-2), 1680 cm^{-1} (Table 5A-3) and 1710 cm^{-1} (Table 5A-4) . 15 °C jumps were performed with an initial temperature from 15 to 45 °C. Data reported includes the observed magnitude and relaxation lifetime of each event.

Table 5A-1: Relaxation Kinetics Probed in the Turn of Loop 1 (1613 cm^{-1}) of FIP35 WW Domain

pH 7						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	1.92 ± 0.02	1.15 ± 0.01	0.70 ± 0.04	94 ± 1	9.0 ± 0.1	800 ± 80
20-35	2.86 ± 0.04	2.91 ± 0.02	0.91 ± 0.02	106 ± 2	17.3 ± 0.2	161 ± 8
25-40	4.79 ± 0.05	2.50 ± 0.04	2.06 ± 0.04	82 ± 1	20.4 ± 0.4	146 ± 5
30-45	4.04 ± 0.04	2.86 ± 0.05	2.05 ± 0.05	96 ± 2	18.7 ± 0.4	110 ± 4
35-50	3.40 ± 0.08	3.6 ± 0.2	2.1 ± 0.2	108 ±4	14.7 ± 0.7	56 ± 6
40-55	3.10 ± 0.08	1.7 ±0.2	6.6 ±0.2	122 ± 9	5.2 ± 0.5	22.7 ± 0.5
45-60	4.2 ± 0.1	2.1 ± 0.2	5.7 ± 0.2	105 ± 4	4.0 ± 0.3	19.0 ± 0.6
pH 5.8						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	0.89 ± 0.03		2.37 ± 0.01	121 ± 6		130 ± 10
20-35	0.86 ± 0.02	0.36 ± 0.01	2.56 ± 0.01	138 ± 7	6.6 ± 0.3	235 ± 2
25-40	0.81 ± 0.03		2.11 ± 0.01	110 ± 8		216 ± 2
30-45			2.20 ± 0.02			241 ± 5
35-50	1.29 ± 0.04		1.70 ± 0.01	106 ± 5		56 ± 7
40-55	1.64 ± 0.05		1.56 ± 0.01	109 ± 5		41.0 ± 0.6
45-60	0.81 ± 0.04		1.23 ± 0.01	120 ± 10		19.8 ± 0.4

Table 5A-2: Relaxation Kinetics Probed in the Sheet of WW Domain (1636 cm^{-1}) of FIP35 WW domain

pH 7						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	2.34 ± 0.04	1.95 ± 0.02	10.84 ± 0.02	93 ± 2	19.1 ± 0.3	242 ± 1
20-35	4.67 ± 0.07	4.55 ± 0.03	8.87 ± 0.03	86 ± 2	18.9 ± 0.2	213± 2
25-40	2.07 ± 0.04	2.97 ± 0.03	11.13 ± 0.02	211 ± 7	13.8 ± 0.2	154 ± 1
30-45	5.01 ± 0.08	5.0 ± 0.1	10.6 ± 0.1	93 ± 2	21.5 ± 0.5	111 ± 2
35-50	4.98 ± 0.06	4.81 ± 0.08	12.98 ± 0.07	95 ± 2	13.9 ± 0.2	70 ± 1
40-55	5.42 ± 0.07	3.2 ± 0.1	11.8 ± 0.1	97 ± 2	7.7 ± 0.3	32 ± 1
45-60	4.18 ± 0.09	2.9 ± 0.2	9.0 ± 0.2	95 ± 3	5.7 ± 0.3	24 ± 1
pH 5.8						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	1.54 ± 0.02	1.66 ± 0.02	5.51 ± 0.02	144 ± 4	37.2 ± 0.7	286 ± 2
20-35	0.90 ± 0.03	1.94 ± 0.07	3.93 ± 0.06	160 ± 10	39 ± 1	196 ± 4
25-40			10.07 ± 0.02			147 ± 1
30-45	1.26 ± 0.06		7.84 ± 0.01	128 ± 9		95 ± 1
35-50	1.21 ± 0.06	3.6 ± 0.1	6.0 ± 0.1	210 ± 20	33 ± 1	153 ± 4
40-55	0.50 ± 0.05	0.7 ± 0.1	5.5 ± 0.1	350 ± 80	8 ± 2	42 ± 1
45-60		1.0 ± 0.1	4.4 ± 0.1		7.9 ± 0.8	28 ± 1

Table 5A-3: Relaxation Kinetics Probed in the Sheet of WW Domain (1680 cm^{-1}) of FIP35 WW domain

pH 7						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	1.41 ± 0.04	0.65 ± 0.02	1.95 ± 0.02	63 ± 3	38 ± 2	403 ± 7
20-35	3.29 ± 0.08	1.82 ± 0.04	1.91 ± 0.04	64 ± 3	22.3 ± 0.7	205± 7
25-40	2.82 ± 0.07	1.55 ± 0.03	2.19 ± 0.03	66 ± 2	24.7 ± 0.7	257 ± 6
30-45	1.70 ± 0.03	2.49 ± 0.04	1.51 ± 0.03	81 ± 3	22.4 ± 0.3	140 ± 5
35-50		2.5 ± 0.2	2.5 ± 0.2		18 ± 1	86 ± 8
40-55	3.86 ± 0.07		6.57 ± 0.01	98 ± 2		17 ± 1
45-60		0.68 ± 0.02	4.60 ± 0.02		2.0 ± 0.1	20 ± 1
pH 5.8						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30	0.23 ± 0.03	0.33 ± 0.01	0.97 ± 0.01	80 ± 20	36 ± 2	440 ±20
20-35	0.32 ± 0.02	0.38 ± 0.02	0.81 ± 0.02	66 ± 8	35 ± 2	227 ± 9
25-40			1.49 ± 0.01			168 ± 2
30-45	1.24 ± 0.04		1.23 ± 0.01	84 ± 4		185 ± 4
35-50			1.52 ± 0.01			75 ± 1
40-55	0.49 ± 0.03		0.9 ± 0.01	160 ± 10		32 ± 1
45-60	0.71 ± 0.03		0.59 ± 0.01	135 ± 9		19 ± 1

Table 5A-4: Relaxation Kinetics Probed in the aspartic acid side chain (1710 cm^{-1}) of Fip35 WW domain

pH 7						
T-jump ΔT(Ti-Tf)/°C	A_1 (mOD)	A_2 (mOD)	A_3 (mOD)	τ_1 (ns)	τ_2 (μs)	τ_3 (μs)
15-30		0.43 ± 0.01			26.1 ± 0.9	
20-35		1.79 ± 0.01			26.8 ± 0.2	
25-40		1.77 ± 0.01			28.0 ± 0.3	
30-45		1.77 ± 0.01			20.9 ± 0.2	
35-50		1.93 ± 0.01			17.5 ± 0.2	
40-55		2.65 ± 0.01			16.7 ± 0.1	
45-60		1.59 ± 0.01			14.2 ± 0.3	

*not enough absorbance change at pH 5.8 to collect T-jump data

Chapter 6: Membrane-Induced Folding of a Cationic Anti-Cancer Peptide

6.1 Abstract

Cationic anti-cancer peptides preferentially disrupt cancer cell membranes. By better understanding this mechanism, we may be able to more effectively target cancer cells. Electrostatic interactions between the positively charged anti-cancer peptide therapeutic and the negatively charged outer leaflet of the cancer cell membrane results in membrane-induced folding of the peptide. SVS-1 is an 18-residue anti-cancer peptide that is designed to be disordered in solution but fold into a β-hairpin at an anionic bilayer surface. A combination of equilibrium circular dichroism and fluorescence measurements were used to characterize SVS-1 association, folding and insertion into the membrane. Secondary structure formation was found to be highly dependent on the peptide: lipid ratio. Temperature dependent measurements reveal that the peptide is folded at the surface of the membrane at all temperatures, but inserts into the membrane only in the fluid phase. These experiments demonstrate how the complex relationship between SVS-1 and the lipid membrane regulate association, folding and insertion into the membrane.

6.2 Introduction

Cancer, caused by the growth and spreading of abnormal cells in an uncontrolled manner, is one of the leading causes of death globally.[145] The current methods that are used to treat cancers, such as surgery and chemotherapy, have a low success rate, risk of reoccurrence, and often have negative side effects.[146] Because of this, much effort in

cancer research has been devoted to designing new cancer therapies. Recently, a class of antimicrobial peptides has been shown to have anti-cancer activity due to electrostatic interactions with the cell membrane.[147] One of the primary differences between normal and cancer mammalian cells is the charge of the cell membrane. Malignant cells are characterized by cell membranes with a net negative charge, whereas, normal mammalian cell membranes are zwitterionic.[148, 149] Antimicrobial peptides are amphiphilic, with opposing cationic and hydrophobic faces. They are thought to interact with the negatively charged cell membrane through the cationic face. Once bound to the cell membrane the hydrophobic face drives insertion of the peptide into the membrane, resulting in cell death.[148-153] Most antimicrobial peptides that are active against cancer cells are α-helices. α-helical peptides that disrupt cancer cells are thought to fold on the surface of the membrane, whereas β-rich peptides are folded in solution.[148-153] Membrane-induced folding of the peptide, like that found in α-helical antimicrobial peptides, offers a higher selectivity for cancer cells, because the peptide must be folded to disrupt the membrane. The SVS-1 peptide is the first β-hairpin to be unfolded in solution and fold at the surface of a negatively charged membrane.[154]

Here we have tested the folding mechanism of SVS-1 using a combination of circular dichroism and fluorescence spectroscopy. SVS-1, KVKVKVKV^{D}P^{L}PTKVKVKVK, is an engineered 18-residue peptide containing N- and C- terminal strands with alternating valine and lysine residues and a type II' β-turn.[154] The turn is made up of the tetrapeptide V^{D}P^{L}PT, which is structured even in the absence of the β-sheets. The peptide was designed to be unfolded in solution and in the presence of neutral charged membrane surfaces. In the presence of a negatively charged membrane surface the peptide binds the

outer leaflet allowing the peptide to fold and insert into the membrane. Circular dichroism was used to demonstrate the selectivity of binding in model membranes made up of neutral and charged large unilamellar vesicles (LUVs). We have employed two approaches to improve our understanding of the folding mechanism. In the first approach, temperature-dependent circular dichroism was used to study the temperature dependence of the secondary structure. In the second approach, fluorescence spectroscopy of a tryptophan mutant was used to probe temperature-dependence of insertion of the peptide into the membrane.

Valine 8 in the turn was selected as the mutation site for tryptophan, because both residues are hydrophobic and mutation of a residue in the sheet may disrupt the amphiphilic structure. We confirmed the selectivity of the SVS-1 W8V mutant by circular dichroism. The phase of the membrane, fluid vs gel or crystalline, can be modulated by the model membrane system. Model membrane systems with phase transitions between -2 and 55 °C were selected to study the dependence of binding, folding and insertion on rigidity of the membrane surface. Changes in secondary structure and tryptophan water solvation with temperature were monitored by circular dichroism. We found that SVS-1 W8V was folded into a β-hairpin in the presence of charged vesicles across all temperatures and model membranes probed. However, subtle differences in the circular dichroism spectra suggest that the peptide adopts a slightly different conformation with vesicles in a fluid phase and vesicles in the gel phase. Fluorescence measurements demonstrate that this difference arises from insertion of SVS-1 W8V in the membrane in the fluid phase. This supports the previously proposed

mechanism that the peptide is unfolded in solution, binds to the electronegative surface of cancer cells, folds and then inserts into the membrane disrupting and killing the cell.

6.3 Experimental Section

Protein Synthesis and Purification. SVS-1 W8V, KVKVKVKW^{D}P^{L}PTKVKVKVK, was synthesized via standard 9-fluorenylmethyloxycarbonyl (Fmoc) base solid-phase chemistry on a Liberty1 microwave peptide-synthesizer (CEM, Matthews, NC). Fmoc-PAL-PS resin (Applied Biosystems, Foster City, CA) was used to form a peptide amide. The peptide was purified by reverse-phase chromatography (C18 column) using a water/acetonitrile gradient with 0.1% trifluoroacetic acid (TFA) as the counter-ion. The peptide was lyophilized and dissolved in a 2 mM HCl solution to allow exchange of the TFA counter-ion for HCl.[90] The identity of the peptide was confirmed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Stock peptide was prepared at 1 mM concentration in water.

Liposome Preparation. Lipids [1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphochoine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol) (DMPG), 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 1,2-dipalmitoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DPPG), 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC), 1,2-distearoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DSPG)] and extrusion kit were purchased from Avanti Polar Lipids, Inc (Alabaster, Alabama). Large unilamellar vesicles (LUV) were used as a model system for CD and fluorescence studies. LUV of 1:1 PC/PG or pure PC was prepared by extrusion

techniques. Liposomes dissolved in chloroform were mixed at the desired molar ratio (1:1 PC/PG or pure PC) and dried by N_2 gas. Solvent removal was completed in a vacuum overnight. The lipid film was reconstituted in 20 mM sodium phosphate buffer and 300 mM NaCl pH 7.4 and subjected to six freeze-thaw cycles. The lipid solutions were extruded using 100-nm pore size filters to obtain LUV. Peptide and lipid were diluted to the desired molar ratio with a final buffer concentration of 10 mM sodium phosphate and 150 mM NaCl pH 7.4.

CD Spectroscopy. CD wavelength scans and CD melting curves were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Jasco temperature controller module (Jasco, Inc., Easton, MD). All measurements were obtained using a 1-mm pathlength cell. Wavelength scans were recorded over the range of 260 to 200 nm with an average of 3 repeats. During spectral acquisition, a bandwidth of 2 nm and scan rate of 50 nm/min was used. Thermal experiments were performed by monitoring the signal at 222 nm from 5 to 95 °C using a 0.1 °C interval and scan rate of 30 °C/hr. During the thermal unfolding experiment a full wavelength scan was obtained every 5 °C after a 60 second delay.

Fluorescence Spectroscopy. Fluorescence spectra were recorded on a HORIBA Jobin Yvon FluoroMax-3 Spectrofluorometer (HORIBA Scientific, Edison, NJ). All measurements were obtained using a 1.0 cm pathlength cuvette. The peptide was excited at 280 nm and monitored from 300 nm to 450 nm. Thermal experiments were monitored every 5 °C from 5 to 90 °C.

Results and Discussion

6.4 Far-UV CD Spectroscopy. SVS-1 W8V and model membranes were used to test the proposed mechanism that the β-hairpin folds at the membrane surface. Far UV-CD was used to investigate the secondary structure of SVS-1 W8V in the presence and absence of model membranes, Figure 6-1. Typically, β-sheet peptides have a CD spectrum with a negative peak at ~218 nm and a positive peak at ~195 nm.[91] Disordered peptides have very low ellipticity above 210 nm and a negative band near 195 nm.[155] The CD spectra of SVS-1 W8V in buffer and in the presence of neutral lipid vesicles exhibit a negative peak at ~198 nm and a positive peak at 223 nm. The negative peak at ~198 nm is consistent with the expected disordered structure. A positive peak at ~230 nm arises from the presence of ordered aromatic side chains.[91] The SVS-1 peptide was designed with a type II' β-turn, $V^{D}P^{L}PT$.[154] In solution, the turn of the SVS-1 peptide is structured, but the β-sheet of peptide is disordered leading to the random coil CD spectrum. The valine in the turn of SVS-1 was mutated to tryptophan in SVS-1 W8V. The CD spectrum indicates that the residue is in an ordered environment, likely packed within the structured turn.

These experiments demonstrate that SVS-1 W8V is unfolded in solution and in the presence of a purely neutral membrane. However, the CD spectrum of SVS-1 W8V in the presence of negatively charged lipid vesicles, made from a 1:1 mixture of DPPC/DPPG, has a minimum at 218 nm and a maximum ~200 nm, consistent with β-sheet structure. This demonstrates that the peptide folds into a β-hairpin in the presence of negatively charged lipids. Folding of the SVS W8V peptide is highly dependent on the peptide:lipid

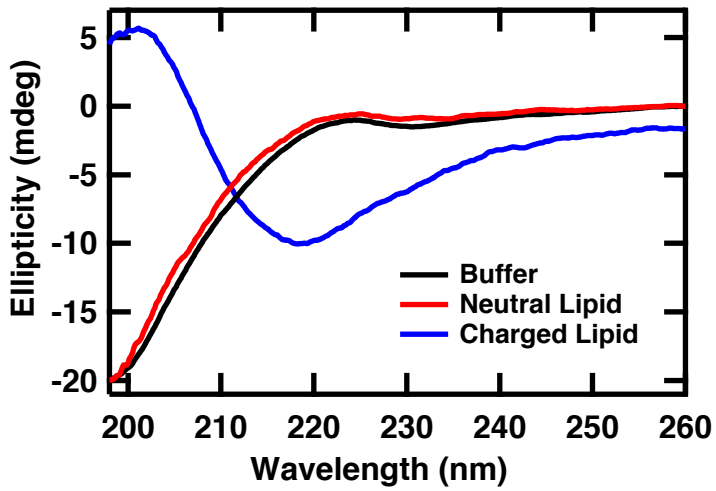


Figure 6-1. Far-UV CD spectra of 60 μM SVS-1 W8V in buffer solution (black, 10 mM sodium phosphate buffer, 150 mM NaCl pH 7.4), neutral 2.5 mM DPPC LUVs (red), and negatively charged 2.5 mM 1:1 DPPC/DPPG LUVs at 20 °C.

ratio (Figure 6-2). Below a peptide:lipid ratio of 1:12 the CD spectra is dominated by a negative peak at 198 nm, consistent with disordered structure. As the peptide:lipid is increased above 1:12 a negative peak at 218 nm and a positive peak at 200 nm appears, consistent with β-hairpin formation. It was not possible to raise the peptide:lipid ratio above 1:100, because of scattering due to precipitant in the sample. The peak associated with an aromatic residue that is in a rigid environment is not present in the folded SVS-1 W8V spectrum (Figure 6-1). This is evidence that the tryptophan undergoes a conformational change upon folding. Tryptophan is a hydrophobic residue. In buffer, the tryptophan likely buries into the ordered turn of SVS-1 W8V, which provides the most protection from solvent. When SVS-1 W8V is folded at the surface of the charged

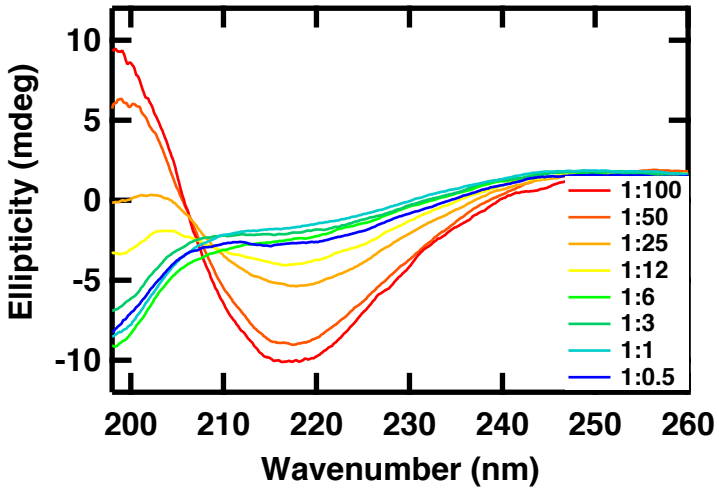


Figure 6-2. Far-UV CD spectra of 50 μM SVS-1 W8V mixed with 1:1 DPPC/DPPG LUVs at peptide:lipid molar ratios between 1:100 and 1:0.5 in 10 mM sodium phosphate buffer, 150 mM NaCl pH 7.4 at 20 °C.

vesicles, the tryptophan buries into the surface of the membrane, which provides greater protection from solvent. The tryptophan buried into the membrane surface is able to freely rotate, leading to the absence of the peak at ~230 nm. Interestingly, this implies that the SVS-1 W8V peptide is not associating with the neutral vesicles. If the peptide were at the surface of the neutral vesicle, the tryptophan would bury into the surface of the membrane and the peak at ~230 nm would disappear.

Temperature dependent experiments were conducted to study the thermal denaturation of the β-hairpin and dependence of the secondary structure on the fluidity of the model membrane. One property of the vesicles is that the mobility of the individual lipid molecules changes with temperature from solid to liquid phase, called gel and fluid phases, respectively. As the temperature is raised the vesicle undergoes two transitions: a

pre-transition from a gel phase to a mixture of gel and fluid phase and a main transition to the fluid phase. We compare the temperature dependence of the SVS-1 W8V β-hairpin in the presence of vesicles that are always in the fluid phase, 1:1 POPC/POPG which has a main phase transition of -2 °C, and in the presence of vesicles that go through a phase transition during the temperature-dependent experiments, 1:1 DSPC/DSPG which has a main phase transition of 55 °C (Figure 6-3).[156] SVS-1 W8V exhibits a negative peak at ~218 nm and a positive peak at 200 nm at all temperatures probed, consistent with a β-hairpin secondary structure. In the fluid phase LUVs (Figure 6-3 A,B), there is no change in the intensity of the peaks over the entire temperature range, demonstrating that the β-hairpin is resistant to thermal changes. However, there is an apparent change in intensity and a shift in the minimum peak from ~218 nm to ~221 nm as vesicles go through a phase transition (Figure 6-3 C,D). This may indicate that the peptide inserts into the vesicle as it changes phases. Inspection of the SVS-1 W8V mixed with 1:1 POPC/POPG LUVs in the fluid phase (Figure 6-3 A) and 1:1 DPPC/DPPG in the gel phase (Figure 6-2) reveals that the minimums are also slightly shifted, 221 nm and 218 nm, respectively. This supports the assignment of the 221 nm minimum to the inserted hairpin and the 218 nm minimum to the hairpin at the membrane surface. This transition (Figure 6-3 D) was fit to an apparent two-state equilibrium model:

$$A_0 = \frac{A_f}{1+exp\left(-\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)\right)} + \frac{A_u}{1+exp\left(\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)\right)} \quad (6\text{-}1),$$

where A_0 is the observed absorbance, A_f and A_u are the absorbance contributions from the folded and unfolded populations, ΔH is the enthalpy change at the midpoint, R is the gas constant, and T_m is the transition midpoint.[134] The transition fits to a midpoint of 53.5 ±

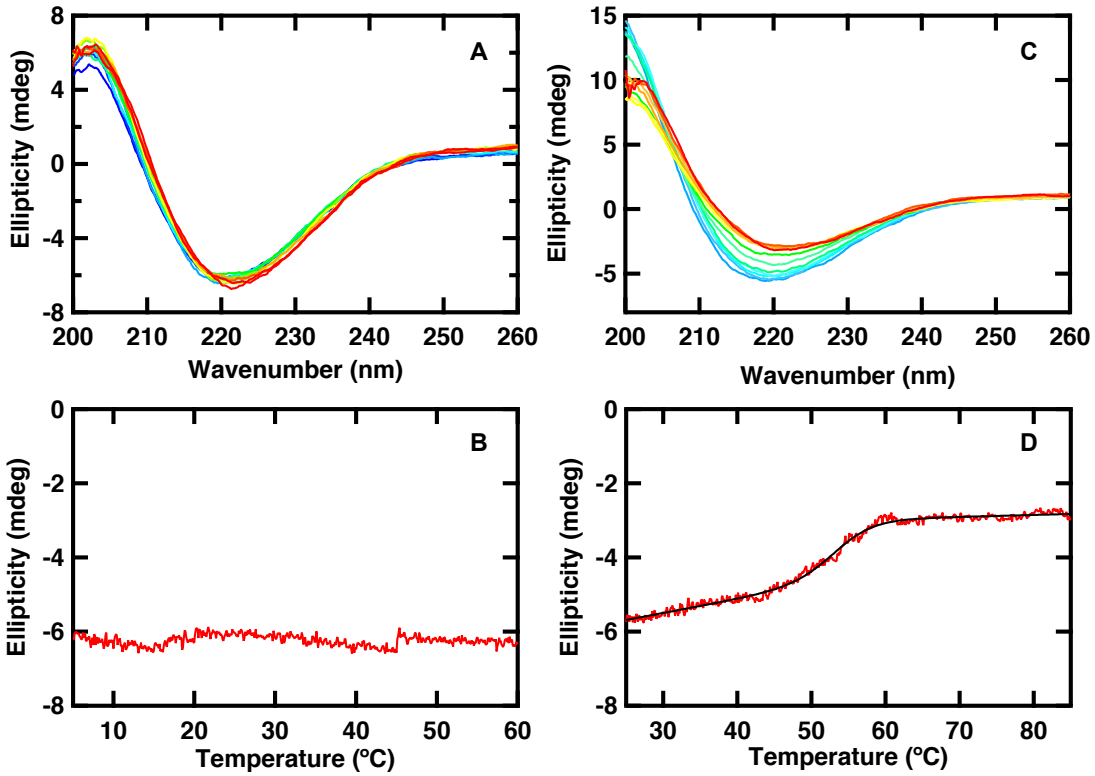


Figure 6-3. Far-UV CD spectra of 60 μM solutions of SVS-1 W8V mixed with 1:1 POPC/POPG LUVs (A, B) and 1:1 DSPC/DSPG LUVs (C,D) at a 1:50 peptide:lipid ratio in 10 mM sodium phosphate buffer, 150 mM NaCl pH 7.4. (A, C) CD spectra acquired every 5 °C during the course of a temperature-dependent experiment in a 0.1 cm pathlength cell. (B, D) Temperature-dependent circular dichroism of the SVS-1 W8V β-hairpin monitored by CD at 218 nm. The continuous line represents the best fit of the data to a sigmoid (Equation 6-1).

0.1 °C. This is close to the transition temperature of the 1:1 DSPC/DSPG LUVs, 55 °C. The transition likely represents a structural change upon insertion that is still largely β-sheet in structure.

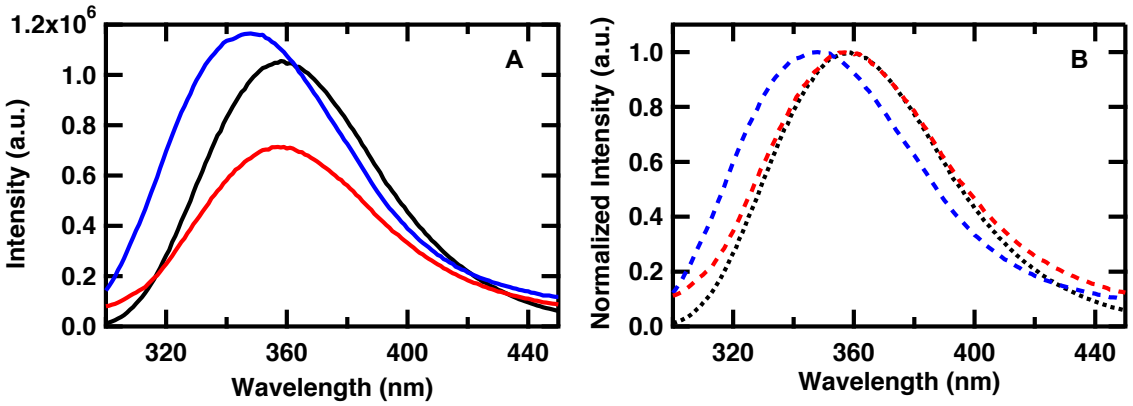


Figure 6-4. (A) Fluorescence spectra and (B) normalized fluorescence spectra of 60 μM SVS-1 W8V in buffer solution (black, 10 mM sodium phosphate buffer, 150 mM NaCl pH 7.4), neutral 2.5 mM POPC LUVs (red), and negatively charged 2.5 mM 1:1 POPC/POPG LUVs excited at 280 nm.

6.5 Fluorescence Spectroscopy. Fluorescence was used to investigate the environment of the tryptophan in the turn of SVS-1 W8V in the presence and absence of model membranes, Figure 6-4. Lipid-peptide interactions can be monitored through changes in the tryptophan fluorescence emission properties of the peptide upon interaction with the model membranes.[157] The fluorescence intensity and wavelength maximum of tryptophan are sensitive to its local environment and have been correlated with the degree of solvent exposure of the chromophore.[158-160] A decrease in intensity and red shift of the maximum wavelength have been reported upon solvent exposure. Model membranes in the fluid phase, 1:1 POPC/POPG, and gel phase, 1:1 DMPC/DMPG (Tm 24 °C) and 1:1 DSPC/DSPG, were used for fluorescence studies.[156] The fluorescence spectrum of SVS-1 W8V in buffer has a wavelength maximum at 360 nm (Figure 6-4 A). First we will analyze the fluorescence spectra of model membranes in the fluid phase. Addition of

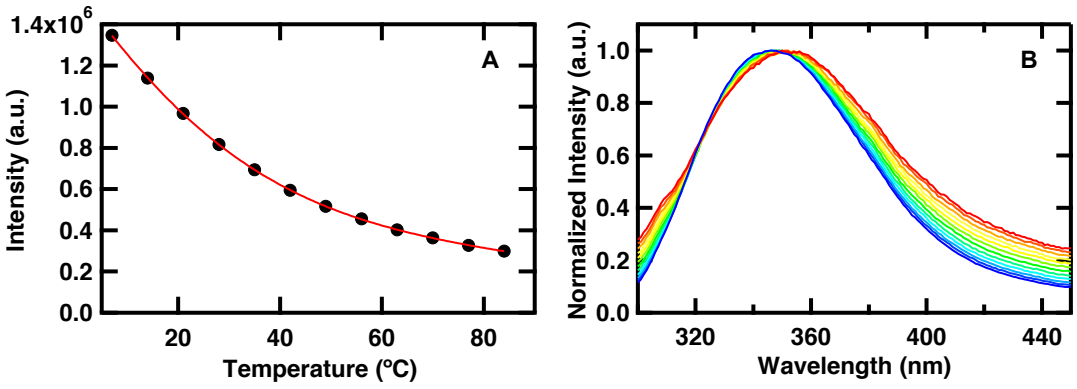


Figure 6-5. Fluorescence spectra of 60 μM SVS-1 W8V in 2.5 mM 1:1 POPC/POPG LUVs in 10 mM sodium phosphate acquired every 5 °C from 5 to 85 °C during the course of a temperature-dependent experiment in a 1 cm pathlength cell. (A) Intensity at the wavelength maximum, 348 nm, monitored by fluorescence excited at 280 nm. The continuous line represents the best fit of the data to a sigmoid (Equation 6-1). (B) Normalized fluorescence spectra excited at 280 nm and normalized at the wavelength maximum.

neutral POPC LUVs results in a decrease in intensity, whereas, addition of charged 1:1 POPC/POPG LUVs results in an increase in overall intensity. Analysis of the wavelength maximum reveals no difference in the peptide in buffer and neutral vesicles, 360 nm, and a blue shift to 348 nm in the presence of negatively charged vesicles. The tryptophan is less solvent exposed in the negatively charged vesicles. This is consistent with CD measurements that show that SVS-1 W8V is folded at the surface of the membrane only in the charged vesicles.

Temperature-dependent fluorescence experiments were conducted to study environmental changes around the tryptophan with temperature. As the temperature is

increased the fluorescence intensity decreases, but with no clear transition (Figure 6-5 A). The temperature dependence of the quantum yield of tryptophan is large, with a fivefold change observed between 5 and 50 °C.[161] Observed changes in the fluorescence intensity are likely due to the intrinsic temperature-dependence of the tryptophan fluorescence. As the temperature is increased there is broadening of the fluorescence spectrum and a small red shift of the maximum wavelength from 348 nm to 351 nm. There is no change in the secondary structure of the protein probed by CD (Figure 6-3 A), so this is not due to a change in conformation. Instead, at high temperatures water may penetrate into the fluid phase of the 1:1 POPC/POPG LUVs causing an apparent shift in tryptophan wavelength maximum.

Model membranes in the gel phase, 1:1 DMPC/DMPG and 1:1 DSPC/DSPG, reveal differences between the interaction of the protein with lipids in the gel and fluid phases. Addition of charged vesicles mixed below T_m results in a wavelength maximum of 353 nm and mixing of charged vesicles above T_m results in a wavelength maximum of 348 nm (Figure 6-6). Here the intensities are not compared because it is not possible to separate the changes due to the environment and changes due to the intrinsic temperature dependence of the quantum yield of tryptophan fluorescence. The blue shift of the wavelength maximum above T_m indicates that the tryptophan is more solvent protected. This likely corresponds to insertion of the peptide into the fluid phase. Below T_m the wavelength maximum is blue shifted relative to the peptide in water. The peptide is likely associated with the membrane surface, providing some solvent protection of the indole ring.

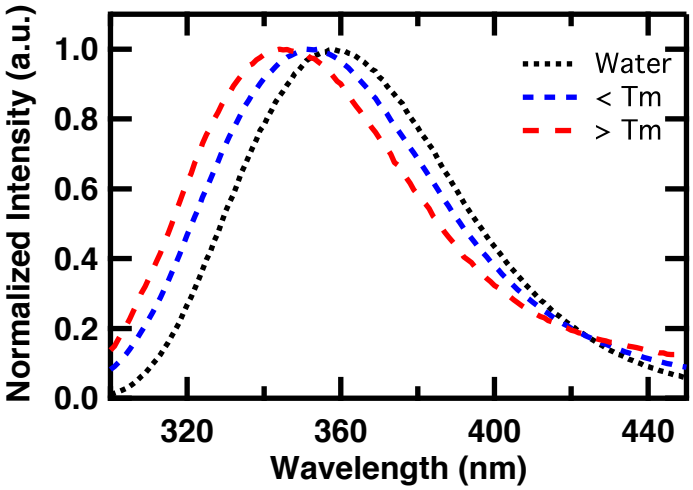


Figure 6-6. Normalized fluorescence spectra of 60 μM SVS-1 W8V in 2.5 mM 1:1 DMPC/DMPG LUVs in 10 mM sodium phosphate buffer, 150 mM NaCl pH 7.4 excited at 280 nm mixed below T_m, 7 °C, and above T_m, 35 °C.

Fluorescence measurements of SVS W8V in the presence of 1:1 DSPC/DSPG were taken every 5 °C over the course of a temperature scan from 5 to 85 °C (Figure 6-7). At low temperatures the fluorescence intensity follows the typical temperature dependence due to the quantum yield of tryptophan (Figure 6-5 A).[161] As the melting temperature is approached there is a large increase in intensity followed by a decrease (Figure 6-7). This transition was fit to an apparent two-state model (Equation 6-1) with a T_m of 50 ± 1 °C. This is similar to the transition observed by CD, 53.5 ± 0.1 °C, and the phase transition of 1:1 DSPC/DSPG LUVs, 55 °C. A large intensity change near the transition of the peptide likely indicates insertion of the peptide. As the temperature is increased there is broadening of the fluorescence spectrum. The shift in wavelength maximum observed in

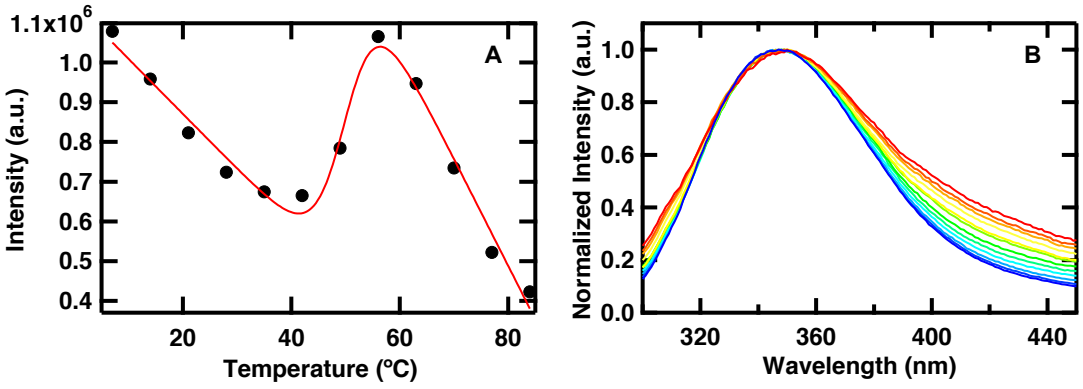


Figure 6-7. Fluorescence spectra of 60 μM SVS-1 W8V in 2.5 mM 1:1 POPC/POPG LUVs in 10 mM sodium phosphate acquired every 5 °C from 5 to 85 °C during the course of a temperature-dependent experiment in a 1 cm pathlength cell. (A) Intensity at the wavelength maximum, 348 nm, monitored by fluorescence excited at 280 nm. The continuous line represents the best fit of the data to a sigmoid (Equation 6-1). (B) Normalized fluorescence spectra excited at 280 nm and normalized at the wavelength maximum.

the lower T_m 1:1 DMPC/DMPG LUVs (Figure 6-6) is not apparent in this data due to the broadening of the fluorescence spectra at high temperatures.

Based on these observations we predict a multi-step model of folding and insertion of anticancer peptide SVS-1 into the cancer cell membrane. The cationic face of SVS-1 W8V interacts with the electronegative surface of the charged membrane. This is supported by CD measurements that demonstrate that not only is the peptide unfolded in the presence of neutral vesicles, but also the structure around the tryptophan is unchanged (Figure 6-1). Tryptophan would likely insert into the membrane surface if the peptide were interacting with the vesicle. Once bound to the outer leaflet of the charged

membrane the peptide is able to fold into a β-hairpin, the bioactive conformation. Comparison of vesicles in the fluid and gel phase show that while the peptide is folded in both cases (Figure 6-3), insertion only occurs when the vesicle is in the fluid phase (Figure 6-5, 6-7). The final step is insertion of the folded β-hairpin into the vesicle.

Most antimicrobial peptides that are active against cancer cells are helical, with only a few examples of peptides with β-structure. The mechanism that helical and β-sheet peptides use against cancer cells is also different. Helical peptides fold in a mechanism similar to that predicted for SVS-1, where they fold into the bioactive conformation at the cell surface.[148-154] β-rich peptides, on the other hand, fold prior to engaging the membrane surface.[148-151, 153] Since folded antimicrobial peptides are capable of disrupting the cell membrane, a peptide that folds only at the surface of electronegative membranes offers higher specificity. Cancer cells have a higher electronegative membrane surface than noncancerous cells, so peptides designed to activate at an electronegative surface will bind and preferentially kill cancer cells.

6.6 Conclusion

SVS-1 is an 18-residue synthetic peptide designed to interact with the negatively charged cell membrane of cancer cells.[154] It is thought to be unstructured in solution, but adopt a β-hairpin structure at the surface of a negatively charged membrane. It is the only anticancer hairpin to date thought to use this mechanism, which is common in α-helical anticancer peptides. Using circular dichroism and fluorescence spectroscopy we confirm that the peptide is folded only at the surface of negatively charged lipids. By modulating the rigidity of the membrane surface we demonstrate that the peptide is always folded at

the membrane surface, but is only able to insert and disrupt a cell membrane in the fluid phase. These experiments confirm the proposed step-wise mechanism for anticancer peptides that the peptide first binds the cell membrane, then folds, finally inserts and disrupts the cell. Characterization of the folding mechanism of such anticancer peptides is necessary for the design of better cancer therapies.

Chapter 7: Conclusion

From this dissertation it is clear that a simple two-state model cannot describe protein folding, but rather folding involves a combination of multiple pathways and intermediate states. We emphasize the importance of using multiple probes, infrared and fluorescence spectroscopy, to resolve these states. These measurements demonstrate that folding is initiated in regions with low contact order, which has been correlated with faster folding times. In addition to contributing to the solution of the protein-folding problem, insights gained from protein folding studies can be applied to the fields of protein misfolding and in the biotechnology industry. These areas require a detailed understanding of the structure and function relationship of proteins.

Proteins cannot function properly in the cell unless they are folded into the correct three-dimensional conformation. There are a number of off pathway routes along the folding pathway from the unstructured protein to its native fold. The cell has developed numerous ways to recover from these misfolded states including using chaperonin proteins to refold misfolded proteins and degradation of misfolded proteins.[162] However, the cell is not always capable of correcting protein misfolding. Accumulation of misfolded proteins has been found in neurodegenerative protein misfolding diseases like Alzheimer's, Huntington's and Parkinson's diseases.[162] Factors such as temperature, pH, concentration and ionic strength all contribute to the susceptibility of proteins to aggregate.[163] Amyloid fibrils, ordered aggregates of β-amyloid peptide, and tangles of protein tau have been implicated in Alzheimer's disease.[164] An estimated 5.3 million American's have Alzheimer's disease in 2015, projected to cost American society $226

billion.[165] The number of American's with Alzheimer's and other neurodegenerative diseases will continue to increase as the population over the age of 65 increases. Protein folding studies reveal fundamental rules of protein folding that will aid in solving the protein-misfolding problem.

Protein folding studies also provide valuable insights that can be utilized in the biotechnology industry: protein engineering and *de novo* design of proteins with novel function.[166] Proteins of pharmaceutical interest, such as insulin, are often overexpressed in bacteria. These proteins can form inclusion bodies, disordered aggregates, which must be refolded to function.[163] Protein engineering may be used to minimize misfolding and increase protein stability for storage. *De novo* protein design can be used to develop new proteins with functions not found in nature. The anticancer peptide SVS-1 studied in Chapter 6 was designed to selectively bind and kill cancer cells. This peptide was engineered based on observations from the folding of antimicrobial peptides.[154] Redesign and design of new proteins based on observations of native protein folding have practical application for pharmaceutical development.

The previous five chapters were split in to three specific aims:

Specific Aim 1: to characterize the folding mechanism and dynamics of a β-hairpin and WW domain *Conclusion: folding of β-sheet structures is heterogeneous*. We took a hierarchal approach to studying β-sheet propagation, starting from the smallest β-sheet structural motif, the β-hairpin, and building up to the WW domain, which incorporates 2 β-hairpins. A combination of Fourier transform infrared spectroscopy and laser induced temperature jumps coupled with infrared and fluorescence spectroscopy were used to probe changes in the peptide backbone. We find that even for the fastest

folding linear β-hairpin, CLN025, folding is heterogeneous and cannot be described by a simple two state model (Chapter 2). Residual interactions in the unfolded state between hydrophobic aromatic residues near the terminus restrain the hairpin leading to a pre-collapsed structure that only requires local rearrangements to reach the folded state. The ultrafast folding rate is a consequence of a negligible global barrier to folding, which also leads to two competitive folding pathways: formation of stabilizing cross-strand hydrophobic interactions and turn alignment. Folding of larger β-systems was also heterogeneous, but not as a consequence of a negligible folding barrier. FBP28 WW domain exhibited multiple relaxation lifetimes covering the nanosecond to millisecond timescale (Chapter 3). These events have not been reported in previous experimental or computational studies. We propose a hierarchical model to describe WW domain folding initiated in the turns, with the β-sheets forming last. One theme that arose from each of these papers is that low contact order, or interactions close in sequence is correlated with faster folding. In the WW domain low contact order is found in the turn, which folds first in our proposed mechanism. In the β-hairpin we observe competition between pathways where the turn or β-sheet forms first; however, there are residual interactions in the termini pinning the end together creating a pseudo-cyclic hairpin with two turns: one in the turn and one at the end of the β-sheet. This pseudo-cyclic hairpin creates a second "turn", creating a second interaction between residues close in sequence space. Therefore a low contact order is found in both the turn and the β-sheet of the β-hairpin. These studies demonstrated the importance of the turn in the folding of β-proteins.

Specific Aim 2: to study the effect of insertion of a fast folding hairpin into larger β-sheet structures *Conclusion: insertion of a fast folding hairpin into larger peptides*

speeds protein folding. What is the role of the hairpin in β-sheet formation? In Specific Aim 1 we observe that residues close in sequence space, the turn of β-proteins, fold on a faster timescale than interactions further apart in sequence space. CLN025 folds on the nanosecond timescale, whereas FBP28 folds on the microsecond timescale. We have systematically replaced each of the loops of FBP28 with CLN025 (Chapter 4). Insertion of the CLN025 β-hairpin into each loop of the FBP28 WW domain reveals that the folding dynamics of the ultrafast folding hairpin is the same within the context of a larger protein. Furthermore, the ultrafast folding hairpin accelerates the overall folding of the WW domain by an order of magnitude when placed in loop 2, since the formation of this loop clearly limits the overall rate of folding. Optimization of the first turn of another WW domain, Pin1, resulted in an increase in the thermal stability and folding kinetics (Chapter 5). Using infrared spectroscopy we observe a protonated side chain in the first turn of WW domains that increases the overall protein stability. Using the methods described in Specific Aim 1 we find that formation of WW domains is hierarchical, initiated in the first hairpin. These studies demonstrate that extended β-sheet structures can be seeded from a hairpin and further than the hairpin stability affects the speed of β-sheet propagation. In each of the faster folding WW domains a residue with a carboxylic acid side chain is found in the turn; electrostatic interactions are important for the formation of β-proteins.

Specific Aim 3: to test the proposed model of hairpin formation and insertion into a membrane *Conclusion: peptide binding to the membrane surface neutralizes peptide charge allowing the hairpin to fold, insert and disrupt the membrane*. Does hairpin folding occur on the surface of the membrane or after insertion? SVS-1, an 18-residue

anti-cancer peptide that is designed to be disordered in solution, but fold into a β-hairpin at an anionic bilayer surface, was used as a model system for protein folding on the membrane (Chapter 6). A combination of equilibrium circular dichroism and fluorescence spectroscopy was used to monitor peptide secondary structure changes and insertion into the hairpin. Electrostatic interactions play an important role in engaging the membrane surface. We observe that the peptide folds on the surface of charged lipids, but inserts only into membranes in the fluid phase. Future kinetic studies will reveal the timescale of folding and insertion into the membrane. Connections between these measurements and folding timescales observed for β-proteins in Specific Aim 1 and 2 will determine the importance of the turn in protein folding *in* vivo. Insight from these studies can be applied towards development of drugs for the treatment of cancer.

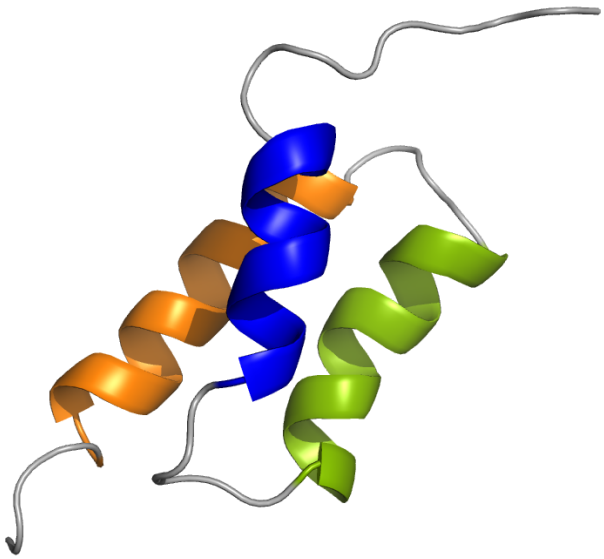
Appendix I: Site-Specific Resolution of Protein Folding using IR Labels Incorporated by Recombinant Protein Expression[167]

A.1 Introduction

While molecular dynamics simulations allow for the resolution of protein folding dynamics at the atomic level, experiments are often unable to resolve folding at even the single-residue level. Structurally specific experimental characterization of the folding dynamics is necessary to test and further refine computational methods. Nonnatural fluorophores and infrared labels can be incorporated into the protein side chains to increase experimental resolution, however, these methods only indirectly report on backbone conformation. Isotope-edited infrared spectroscopy offers a method of site-specifically resolving protein backbone structure and dynamics.[168-170] The amide-I mode of the peptide backbone is sensitive to secondary structure; characteristic infrared bands can be assigned to solvated and buried alpha helix, beta sheet, turns and random coil structures. Isotopic substitution of the carbonyl carbon and carbonyl oxygen shifts the amide I band of the residue of interest to a less congested spectral region. By placing $^{13}C=^{18}O$ labels at different positions in the protein, it should be possible to determine the folding dynamics at any position of interest. The native fold of a protein is comprised of

multiple secondary structure elements stabilized by non-local interactions including the hydrophobic core, salt bridges and hydrogen bonding. Using isotope labels in each of the secondary structure elements it should be possible to determine not only the order of formation of secondary structure elements, but also the dynamics of the labeled residue.

Here we have compared the folding dynamics of the three helices in an ultrafast folder, the B-domain of staphylococcal protein A (BdpA) (Figure A-1), using site-specific isotope labeling. Because of its fast folding rate and simple structure, BdpA has been the focus of extensive experimental and computational studies. While experiment and simulation agree that BdpA folds through an intermediate, there are several conflicting models of the transition state structure.[171-185] Experimental studies support a model of the transition state that includes a fully formed Helix 2, with a partially formed Helix 1 and stabilizing interactions between Helix 1 and Helix 2.[171-176] Computational studies have predicted slightly different models which include either complete formation of Helix 1[177-179] or Helix 3[180-185]. Discrepancies between experimental and computational results may arise from differences in the sequences studied, resolution of experimental probes used or biases in the computational method. In particular, attempts to resolve dynamics at the single-residue level, time-resolved fluorescence and FRET, have relied on incorporation of fluorescent probes, which are not necessarily conservative mutations of the native BdpA sequence and can be difficult to interpret.[174, 175, 186-188] In our approach, $^{13}C=^{18}O$ labeled methionine infrared probes were introduced into each of the helices to independently monitor the dynamics of the backbone of each helix.



TADNKFNKEQQNAFYEILHLPNLNEEQRNG
--------10--------20--------30
FIQSLKDDPSQSANLLAEAKKLNDAQAPKA
--------40--------50--------60

Figure A-1. Top: Cartoon of BdpA (PDB entry: 1BDD) showing Helix 1 (blue), Helix 2 (green) and Helix 3 (orange). Below: The WT BdpA sequence with Helix 1, Helix 2 and Helix 3 colored to match the cartoon. Positions for methionine mutations are underlined. The figure was prepared in PyMOL (www.pymol.org).

An extensive library of point mutations of BdpA has been created for phi-value analysis determination of the transition state structure.[171, 172] In these studies, a Y15W mutation to Helix 1 was utilized as a pseudo-parent protein to provide an optical probe for the phi-value analysis. Mutation of I32 in Helix 2 to valine or alanine resulted in an ~20 °C decrease in melting temperature.[171] This is likely because I32 is found in the

hydrophobic core of BdpA, and mutation disrupts stabilizing interactions between Helix 2 and Helix 1/Helix 3. Phi-value analysis predicted that the N-terminus of Helix 3, probed at residue A47, was formed in the transition state but not the C-terminus, K51.[171] We selected Y15, I32 and A47 as positions for our labels, because of their respective locations in Helix 1-3 and the prediction from phi-value analysis that these positions are in native-like conformations in the transition state. Methionine has a moderately sized side chain, similar to the side chains of tyrosine, isoleucine and alanine and we have shown it does not significantly perturb the structure of BdpA. Therefore, we investigated labeled mutants where the Y15, I32 or A47 is replaced with $^{13}C=^{18}O$ labeled methionine.

We compared the stability and folding kinetics of the wildtype and mutated BdpA. The stability was determined by equilibrium circular dichroism and FTIR measurements. The kinetics was measured using temperature jump coupled with time resolved infrared spectroscopy. Pulsed laser excitation was used to rapidly initiate a shift in the folding equilibrium. The relaxation kinetics of the solvent exposed helix, buried helix, and labeled helix positions were independently measured by probing the components of the IR amide I band assigned to each structure. As expected, the I32M mutation resulted in destabilization of the protein due to disruption of the hydrophobic core. Mutation of Y15 or A47 had little effect on the protein's stability. Kinetics of the solvated and buried helix positions of the mutants agree well with those of the parent BdpA. The IR detected kinetics and melting temperatures of the labeled positions more closely resembles the solvated helix than the buried helix. All of the mutants exhibit two well-separated kinetics phases when probed by infrared spectroscopy, which can be described by a 3 state model where folding proceeds through an intermediate state (Figure A-2). The

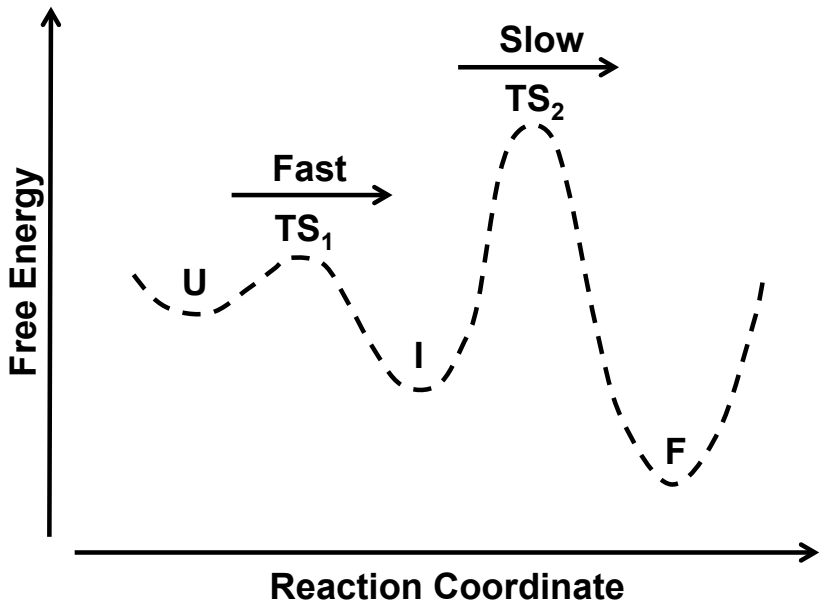


Figure A-2. Folding free energy profile of BdpA and mutants. TS_1 and TS_2 represent transition states between unfolded (U), intermediate (I) and folded (F) states.

second phase is slow and represents crossing the "global" transition state (TS_2) between intermediate and folded states. Interactions present in the intermediate structure will also be present in TS_2, because the intermediate occurs prior to TS_2 on the reaction coordinate. Contacts at the Y15, I32 and A47 positions are all formed in the fast initial phase, an intermediate structure consistent with the "global" transition state predicted by phi-value analysis.[171, 172]

Wavelength dependent infrared measurements at the labeled position reveal a fast 100 nanosecond phase in addition to a slower microsecond phase. These measurements provide evidence for at least partial formation of Helix 1, Helix 2 and Helix 3 in the transition state. Time resolved measurements have previously been unable to resolve a

fast phase in Helix 1 or Helix 3. The increased sensitivity of the infrared probes to local structure has aided in the resolution of early formation of helical structure in Helix 1 and 3, supporting folding models proposed by computer simulations. These results demonstrate the importance of developing experimental probes of protein folding at the single-residue level in comparing experiments to all-atom molecular dynamics simulations.

A.2 Experimental Section

Protein Synthesis and Purification. The BdpA gene was incorporated into a pET-30a(+) vector with a histidine affinity tag and TEVprotease cleavage site at the N terminus. A quikchange kit was used to generate the mutant BdpA plasmid (Agilent Technologies, Santa Clara, CA). Primer oligonucleotides used to introduce the mutations were obtained from Invitrogen (Life Technologies, Grand Island, NY): Y15M 5'-ATTTAACAAAGAACAACAAAACGCCTTTATGGAAATCCTCCACTTACCTAATTTAAA CG -3', I32M 5'-GAAGAACAACGCAATGGTTTTATGCAATCACTCAAAGACGAC-3', A47M 5'-ACGACCCATCTCAATCAGCCAACTTACTTATGGAAGCCAAAAAATTAAACG-3'. Mutated plasmid DNA was recovered using a QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) and the sequence was verified by sending the plasmid for sequencing at Beckman Genomic Center (Danvers, MA). The plasmids containing wild type and mutated BdpA were introduced to the B834pLysS auxotrophic expression cell line using heat shock (Novagen, Foster City, CA). Cultures were grown in kanamycin and

chloramphenicol containing media overnight. Glycerol was added to a final concentration of 15% v/v and glycerol stocks were stored at -80 °C.

For mutant peptides, an overnight culture of B834pLysS was grown in LB rich media with kanamycin and chloramphenicol. The starter culture was added to 1 liter of M9 minimal media supplemented with 40 mg of each amino acid, including unlabeled methionine. Mutant cultures were grown to an OD of 0.8 and then harvested again using centrifugation. Cells were resuspended in minimal media containing the same components, but with 40 mg $^{13}C=^{18}O$ labeled methionine instead of unlabeled methionine. The procedure for $^{13}C=^{18}O$ labeling methionine has been described elsewhere.[189] Protein expression was induced with 1 mM IPTG. (Wildtype cultures in LB media were grown to an OD of 0.8 then induced with 1 mM IPTG.) After 24 hours of growth the cells were harvested by centrifugation. Cells were lysed by stirring and sonication and cell debris were removed by centrifugation. Purification was carried out by FPLC on a Ni His prep FF 16/10 column (GE Healthcare Life Sciences, Pittsburgh, PA). The yield for mutant peptides was 10 mg per liter of expression. Expression was confirmed by SDS PAGE and mass spectrometry.

Isolated proteins containing the histidine purification tag were subjected to a cleavage protocol using AcTEV protease (Invitrogen). Cleaved protein was separated from free tag using a Ni Sepharose High Performance resin (GE Helathcare Life Sciences, Uppsala, Sweden) in a batch protocol. A mixture of resin and protein were mixed overnight on a rotisserie and the supernatant harvested the next day. Clean protein was lyophilized overnight and dissolved in D_2O to allow deuterium-hydrogen exchange of the amide protons. The protein was lyophilized a second time and stored at -20 °C prior to use.

CD Spectroscopy. CD spectra and CD melting curves were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S Jasco temperature controller module (Jasco, Inc, Easton, MD). Peptides were dissolved at ~30 μM in 25 mM potassium phosphate and 50 mM NaCl at pH 6.8. All measurements were obtained using a 1-mm pathlength cell. Wavelength scans were recorded over the range of 260 nm to 190 nm with an average of 3 repeats. A bandwidth of 2 nm, response of 2 seconds, and a scan rate of 100 nm/min were used for spectral acquisition. Thermal unfolding experiments were performed by monitoring the signal at 222 nm from 20 to 100 °C using a 1 °C interval and a ramp rate of 40 °C/hr. After the melt, the temperature was returned to 20°C and a wavelength scan was obtained to determine the reversibility of folding.

FTIR Spectroscopy. The FTIR absorption spectra and equilibrium melting behavior were monitored on a Varian Excalibur 3100 FTIR spectrometer (Varian Inc., Palo Alto, USA) using a temperature controlled IR cell. The IR cell consists of two CaF_2 windows separated by a 130 μm Teflon spacer split into two compartments, a sample and a reference. The same cells are used for equilibrium FTIR and T-jump experiments. Sample concentrations of 0.5-10 mg/ml were prepared for IR experiments. No aggregation was observed in the infrared at the reported concentrations. The infrared spectrum is highly sensitive to protein aggregation since the amide I band of an aggregated protein is dominated by an intense, sharp band at ~1620 cm^{-1} and a coupled peak that is also sharp but not as intense at ~1680 cm^{-1}, due to dipolar coupling of the C=O groups in protein aggregates.[190] Furthermore, the T-jump experiments are also very sensitive to aggregation, because protein aggregates serve as nucleation sites for cavitation, leading to large cavitation artifacts in the IR transients, which were not

observed in our experiments. All spectra shown at a specific temperature are constructed by subtracting the spectrum of reference buffer solution without protein from sample solution with protein. The temperature dependent difference spectra were then generated by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures. The second derivative spectra were computed in IGOR PRO after smoothing the data with a sixth order binomial algorithm to remove any residual water vapor peaks (WaveMetrics, Lake Oswego, OR).

Time Resolved Temperature Jump (T-jump) Relaxation Measurements. The IR T-jump apparatus has been described previously.[30] Pulsed laser excitation is used to rapidly perturb the folding equilibrium on a timescale faster than the molecular dynamics of interest. Time resolved infrared is then used to probe the reaction. The Q-switched GCR-10 Nd:YAG laser (Spectra Physics, Mountainview, CA) fundamental at 1064 nm is Raman shifted (one stokes shift in 200 psi H_2 gas) to produce a 10 ns pulse at 2 μm. The magnitude of the T-jump is calculated using the change in reference absorbance with temperature. The T-jump reference is taken from D_2O buffer with 25 mM potassium phosphate and 50 mM NaCl at pD* 6.8 at the same temperature and probe laser frequency as the sample. Absorbance changes at the reference frequency are due only to changes in D_2O absorption, which is used as an internal thermometer.[30]

The change in signal induced by the T-jump is probed in real time by a continuous laser with a frequency in the amide I' band of the IR. The mid-IR probe beam is generated by a continuous wave quantum cascade laser (Daylight Solutions Inc., San Diego, CA) with a tunable output range of 1570-1730 cm^{-1}. The transient transmission of the probe beam through the sample is measured using a fast, 100 MHz, photovoltaic

MCT IR detector/preamplifier (Kolmar Technologies, Newburyport, MA). Transient signals are digitized and signal averaged (1000 shots) using a Tektronics digitizer (7612D, Beaverton, OR). Instrument control and data collection are controlled using a LabVIEW computer program.

Analysis of Kinetics Data. The peptide relaxation kinetics must be deconvolved from the observed kinetics. Accurate deconvolution is possible as the instrument response is determined from the reference measurement under the exact conditions of the sample measurements. In order to minimize detector artifacts, the reference is scaled prior to subtraction from the sample. The decay function is a double exponential decay with the formula:

$$A = A_0 + A_1 \exp\left(\frac{-(x - x_0)}{\tau_1}\right) + A_2 \exp\left(\frac{-(x - x_0)}{\tau_2}\right) \quad \text{(A-1)}$$

where A_0 is an offset, A_1 and A_2 are preexponential factors, τ_1 and τ_2 are relaxation lifetimes of the sample and x_0 is the time offset. The data are fit over the interval from 95 ns to 200 μs. The data analysis was performed in IGOR PRO (Wavemetrics).

Results and Discussion

A.3 Far-UV CD Spectroscopy. $^{13}C=^{18}O$ labeled methionine mutations were made at single positions in each of the helices of BdpA to determine the folding dynamics of each of the helices independently. Far-UV CD spectra obtained from the wild type (WT) BdpA and $^{13}C=^{18}O$ labeled Y15M (Helix 1), I32M (Helix 2) and A47M (Helix 3) BdpA mutants were almost identical (Figure A-3A), suggesting that the methionine mutations do not significantly perturb the native fold. Thermal denaturation was monitored by

recording the absorbance change at 222 nm with temperature (Figure A-3B). The melting curves were fit to an apparent two-state equilibrium model:

$$A_O = \frac{A_F}{1 + exp\left(-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_M}\right)\right)} + \frac{A_U}{1 + exp\left(\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_M}\right)\right)} \qquad (A-2)$$

where A_O is the observed absorbance, A_F and A_U are the absorbance contributions from the folded and unfolded populations, ΔH is the enthalpy change at the midpoint, R is the gas constant, and T_M is the transition midpoint.[91] The data were then normalized for comparison. The observed melting temperature of wildtype BdpA is 72.7 ± 0.1 °C (Table A-1). This agrees with the previously reported melting temperature of BdpA, 72.5 ± 0.1 °C, also obtained by a thermal melt monitored by far-UV CD of the 222 nm band.[175] Each mutant unfolds sigmoidally with similar unfolding cooperativity to WT BdpA (Table A-1). The Y15M and A47M mutations result in a change in the melting temperature of less than 1°C, suggesting that there is little change to the protein stability. This is further evidence that the methionine mutations do not affect the fold. The I32M mutation results

Table A-1: Summary of Melting Temperatures

	222 nm (°C)	**1648 cm^{-1} (°C)**	**1632 cm^{-1} (°C)**	**1560 cm^{-1} (°C)**
Wild type	72.7 ± 0.1	73.6 ± 0.2	72.0 ± 0.4	n/a
Y15M	71.9 ± 0.1	70.7 ± 0.2	67.9 ± 0.3	67 ± 2
I32M	64.2 ± 0.1	65.5 ± 0.2	64.6 ± 0.5	56 ± 3
A47M	73.3 ± 0.1	74.5 ± 0.1	73.7 ± 0.3	68 ± 1

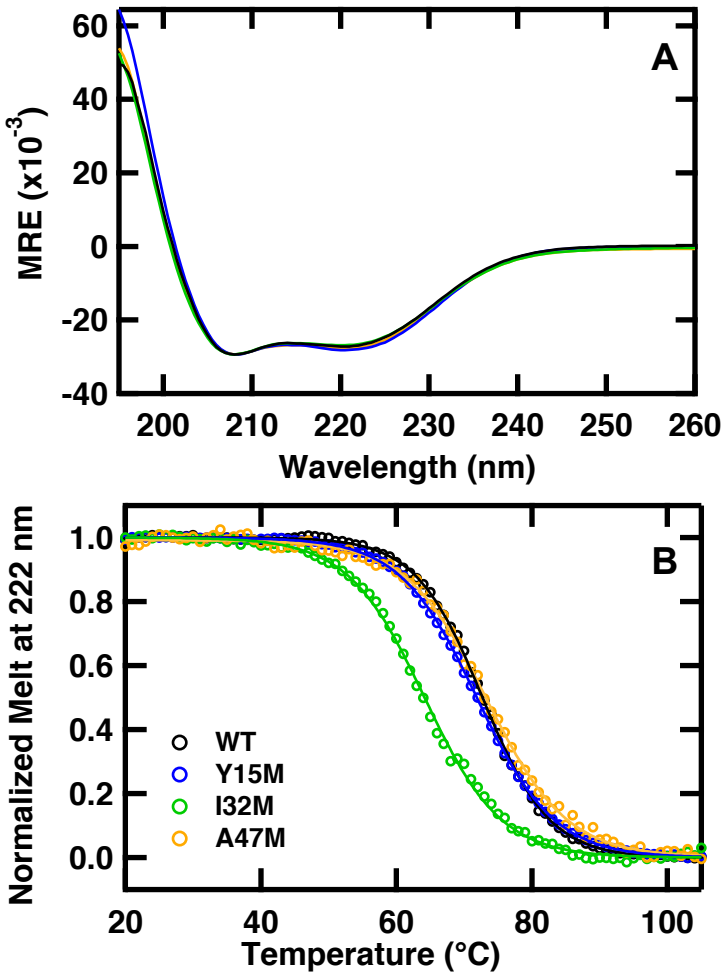


Figure A-3. (A) Far-UV CD spectra of 30 μM solutions of WT (black), Y15M (blue), I32M (green), and A47M (orange) BdpA in 25 mM potassium phosphate and 50 mM NaCl at pH 6.8 acquired at 23 °C. The mutants are normalized to the minimum of WT BdpA. (B) Thermal denaturation of WT BdpA and mutants monitored by CD at 222 nm. The continuous lines are fit to an apparent 2-state model (Equation A-2) and then normalized.

in a decrease in the melting temperature of ~9 °C (Table A-1). I32 is part of the hydrophobic core of BdpA and forms tertiary interactions between Helix 2 and Helix 1/Helix 3.[171] The methionine mutation likely disrupts these contacts leading to a decreased melting temperature. Nevertheless, the CD and IR spectra indicate that I32M still adopts the same folded structure as the WT.

A.4 FTIR Spectroscopy. Comparison of the amide I' (amide I region of peptides in D_2O) FTIR absorption spectra of the WT and mutant proteins shows that they are nearly identical (Figure A-4A, C, plus additional mutants in Appendix Figure AA-1). The amide I' absorbance arises from C=O stretching vibrations of the polypeptide backbone carbonyls, and is an established indicator of secondary structure.[62-64] This relatively broad band contains contributions from the entire polypeptide backbone, which in the case of BdpA includes solvated α-helix, buried α-helix, turns and random coil structure. The FTIR spectra indicate that the secondary structure content is not perturbed by the mutations, consistent with the CD results. The temperature-induced unfolding of WT BdpA and $^{13}C=^{18}O$ labeled methionine mutants was studied over the range from 25 to 100 °C in 5 °C intervals. The temperature dependent absorption spectra of the amide I' spectral region of WT BdpA and BdpA Y15M mutant are shown in Figure A-4A and A-4C. The changes with temperature are highlighted by the difference spectra for each peptide (Figure A-4B and A-4D). The difference spectra are generated by subtracting the lowest temperature spectrum from each absorption spectrum at higher temperature. There are three main components of the amide I' band centered at 1632, 1648 and 1670 cm^{-1}. These peaks have been identified previously in α-helical peptides.[191] Positive peaks correspond to new interactions with solvent in the unfolded state; the characteristic

feature due to the disordered polypeptide is a relatively broad peak, located at 1670 cm^{-1}. Negative peaks correspond to specific structures or interactions present in the folded state. The peak at 1632 cm^{-1} corresponds to solvated helix (the backbone positions along each helix that are exposed to solvent) and the peak at 1648 cm^{-1} corresponds to buried helix (the backbone positions protected from solvent in the hydrophobic interior of the structure). Inter-residue coupling of the C=O stretches affects the frequency and intensity of each of these components, particularly for long helices.[192] A $^{13}C=^{18}O$ label placed in the middle of a helix disrupts this coupling (the coupling through the labeled position is inefficient because of the difference in C=O stretching frequency between labeled and unlabeled groups) leading to differences in intensity of the *unlabeled* 1632 and 1648 cm^{-1} bands of the WT and labeled BdpA (Figure A-4B and A-4D).

The $^{13}C=^{18}O$ labeled methionine is shifted to lower frequency relative to the unlabeled $^{12}C=^{16}O$. The expected shift is 75 cm^{-1} based on the local harmonic oscillator model of the C=O stretching vibration.[189] This peak can be difficult to identify because of its relatively small intensity, overlap with the absorbance of carboxylic acid side chains, and mismatch in HOD content between the sample and reference. The apparent intensity of the labeled peak in Figure A-4D is larger than expected, partially due to mismatch in the HOD content of the sample and reference. The second derivative of the FTIR difference spectra between the highest and lowest temperature can be used to eliminate broad contributions to the baseline and better resolve peaks contributing to secondary structure (Figure A-5). The positive peaks highlight interactions that are lost at high temperature, the 1632 and 1648 cm^{-1} peaks corresponding to loss of solvated and buried helices. The I32M and A47M spectra exhibit a temperature dependent peak at 1586 cm^{-1}, consistent with side

chains of carboxylic acid. This peak arises from residual EDTA from the cleavage buffer solution. This contaminant was removed from the Y15M and WT BdpA samples. Comparison of purified samples to samples with residual EDTA demonstrates that the residual EDTA does not affect the folding or stability of BdpA (Appendix Figure AA-2, AA-3). The $^{13}C=^{18}O$ labeled methionine samples exhibit a peak at 1560 cm^{-1} that is not present in the WT BdpA sample. Comparison of the His-tagged unlabeled methionine mutant to the WT BdpA and His-tagged $^{13}C=^{18}O$ labeled methionine mutant demonstrates

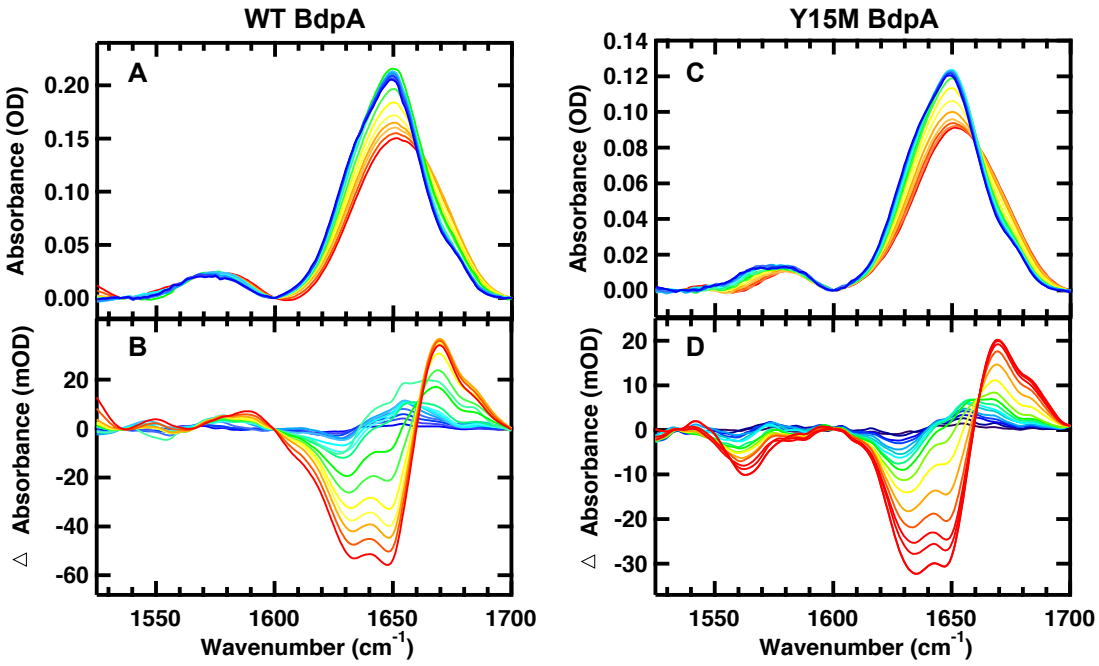


Figure A-4. Temperature dependent FTIR spectra of 6 mg/ml WT BdpA (A, B) and 3 mg/ml Y15M BdpA (C, D) in 25 mM potassium phosphate and 50 mM NaCl (pH 6.8). (A, C) Absorption spectra in the amide I' region: the temperature of the individual traces varies from 25 to 100 °C in 5 °C intervals. (B, D) Difference spectra obtained by subtracting the spectrum at 25 °C from the spectra at higher temperatures.

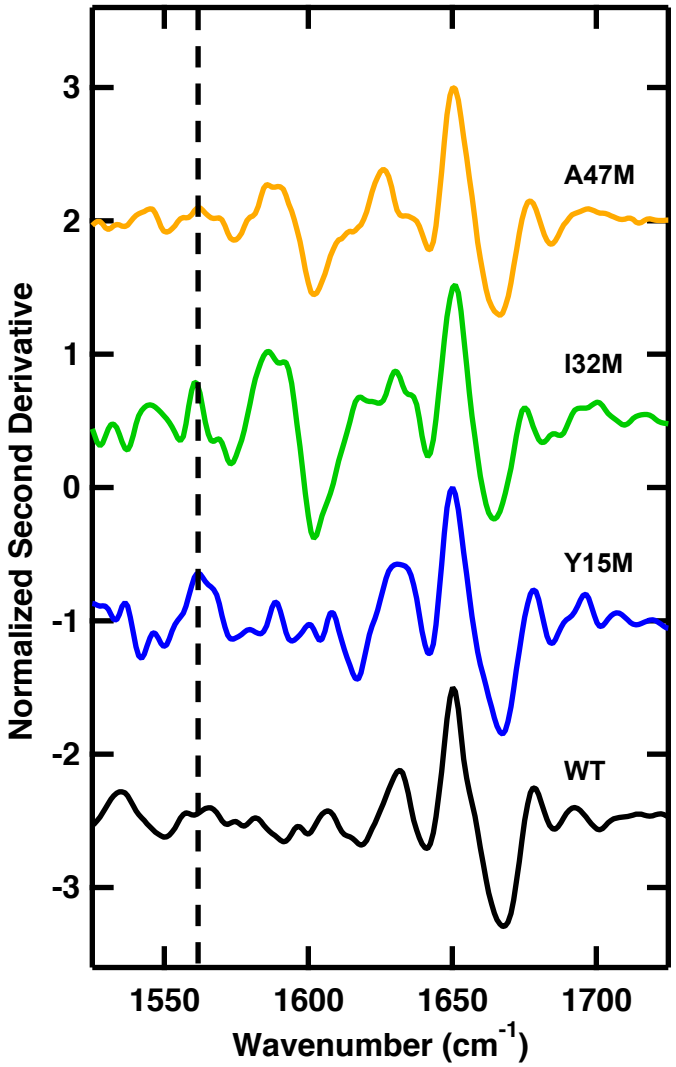


Figure A-5. Second derivative of the FTIR difference spectrum (100-25 °C) of WT BdpA (black), Y15M BdpA (blue), I32M BdpA (green) and A47M BdpA (yellow). The data are normalized at the maximum and offset for clarity. The dashed vertical line at 1560 cm^{-1} highlights the $^{13}C=^{18}O$ labeled peak.

position consistent with an α-helix. The observed shift is 88 cm^{-1} relative to the unlabeled buried helix peak at 1648 cm^{-1} and 72 cm^{-1} relative to the unlabeled solvated helix

that this peak does not arise from differences in side chain packing between the mutant and WT peptide (Appendix Figure AA-4, AA-5). The frequency of the labeled position is determined by secondary structure, with contributions from hydrogen bonding and dipole coupling modifying the amide I position.[40] The strength of hydrogen bonding between the $^{13}C=^{18}O$ and neighboring $^{12}C=^{16}O$ groups will lead to a shift in the labeled amide I position. This is slightly higher or lower than the expected shift of 75 cm^{-1}, suggesting that the $^{13}C=^{18}O$ labeled positions are partially solvated. Partially solvated helices have been demonstrated to have an IR frequency in between a fully solvated, 1632 cm^{-1}, or fully buried, 1648 cm^{-1}, helix frequency.[193] This assignment is also consistent with changes in the ratio of intensities of the *unlabeled* 1632 and 1648 cm^{-1} bands of the labeled proteins. Disruption of the C=O coupling of the solvated position, particularly for the longer helices 2 and 3, results in a decrease in the 1632 cm^{-1} band (solvated helix) intensity relative to the 1648 cm^{-1} band (buried helix) intensity when compared to WT BdpA (Figure A-1S). Consistent with previous experiments of α-helices with a single $^{13}C=^{16}O$ labeled carbonyl, we observe no shift of the $^{12}C=^{16}O$ amide I modes.[194] While incorporation of the $^{13}C=^{18}O$ labels disrupts the coupling of the amide I mode (Figure A-4B and A-4D), such labels can still be described as nonperturbing because the coupling energy is small (a few cm^{-1}) and therefore does not affect the folded structure or thermodynamics of folding.

The normalized melting curves for wildtype BdpA and $^{13}C=^{18}O$ labeled mutants derived from temperature dependent IR absorption are shown in Figure A-6. Infrared melts were obtained at wavelengths corresponding to buried helix (1648 cm^{-1}), solvated helix (1632 cm^{-1}), and the $^{13}C=^{18}O$ labeled position (1560 cm^{-1}). The data were

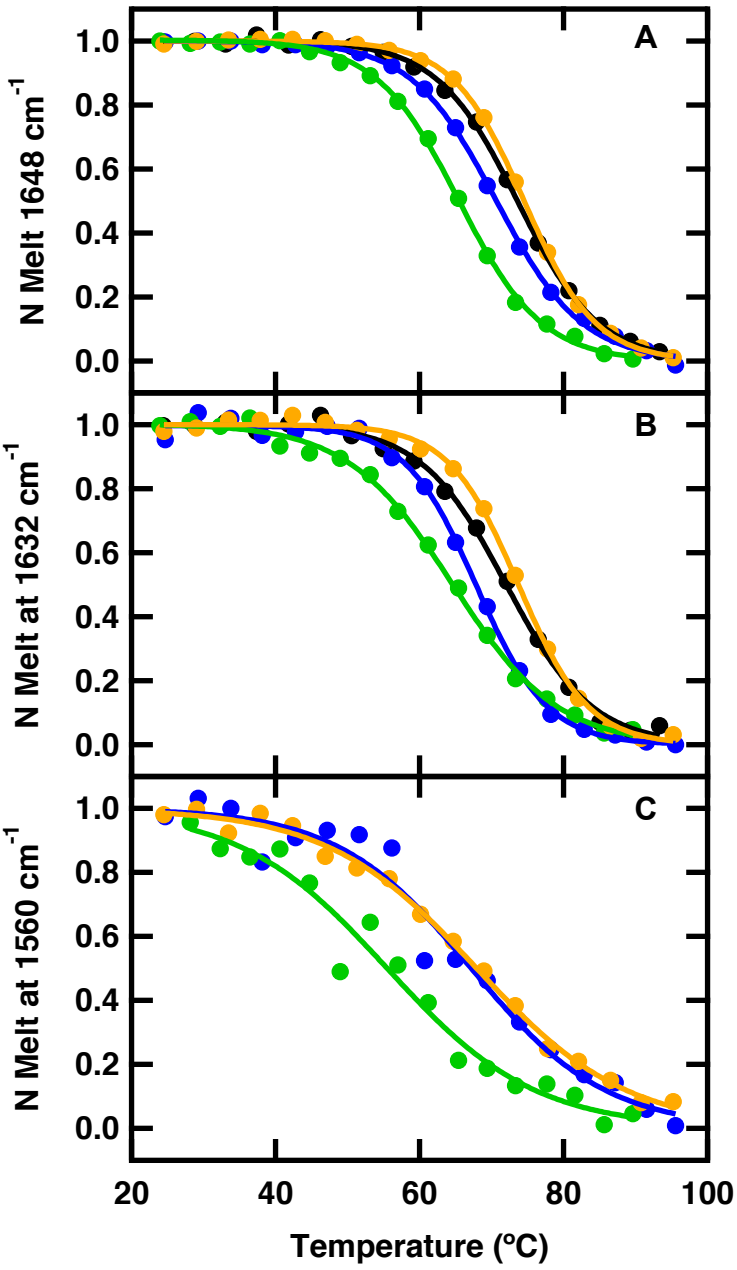


Figure A-6. FTIR melt curves for WT BdpA (black), Y15M BdpA (blue), I32M BdpA (green) and A47M BdpA (yellow) obtained by plotting the change in IR difference spectra at 1648 cm^{-1} (A), 1632 cm^{-1} (B), and 1560 cm^{-1} (C) versus temperature. The data are fit to an apparent 2-state model (Equation A-2) and then normalized.

normalized after being fit to an apparent two-state equilibrium model using Equation A-2. There is good agreement between the melting temperature and breadth of transition probed at 222 nm by circular dichroism and melting temperatures probed at 1632 and 1648 cm^{-1} by FTIR (Figure A-6, Table A-1). The observation of a lower melting temperature of the I32M mutant compared to WT BdpA supports the conclusion from the CD measurements that the I32M mutation destabilizes the hydrophobic core. In all four peptides there is a slight difference in the melting temperature measured at the buried and solvated helix position. The lower melting temperature of the solvated helix band is likely due to fraying of solvent exposed regions of the individual helices prior to the global unfolding transition, which broadens and shifts the apparent transition midpoint to a lower temperature. This is evidence for a folding intermediate with a reduced helical content, lacking some of the less stable helical regions. The $^{13}C=^{18}O$ labeled positions have melting temperatures similar to that of the solvated helix positions, supporting the assignment of this peak as partially solvated. The overall signal change at the $^{13}C=^{18}O$ position is smaller than in the amide I' region leading to the relative increase in noise compared to the unlabeled position. The breadth of the transition at the labeled positions is twice as large as the breadth of the transition at the unlabeled positions. The broader transition is a result of $^{13}C=^{18}O$ position being a local probe, primarily sensitive to local changes in structure and solvation and only indirectly sensitive to the global, cooperative folding transition.

A.5 Temperature-Jump Relaxation Kinetics. The relaxation kinetics of the folding/unfolding transition following a laser induced temperature jump were probed using time resolved infrared spectroscopy of the amide I' frequency of the buried helix

(1648 cm^{-1}), solvated helix (1632 cm^{-1}) and $^{13}C=^{18}O$ labeled methionine (1560 cm^{-1}). Jumps were performed slightly off peak center (near the maxima of the difference features) to maximize the transient absorbance signal. Because the amide I subcomponents are relatively broad (Figure A-4), the off center probe frequencies employed still correspond to the respective structural assignments. Figure A-7 displays the wavelength dependent relaxation kinetics for the I32M mutant following a jump from 50 to 60 °C. All of the relaxation transients are fit by a double exponential, and the lifetimes are reported in Table A-2 (Appendix Figure AA-6, AA-7, AA-8). There is good agreement between the kinetics observed for WT BdpA and the mutants at the unlabeled probe frequencies of 1648 and 1632 cm^{-1} (Table A-2). All exhibit a fast ~100 ns phase (τ_1) and a slower ~10 μs phase (τ_2). These kinetics agree with previous WT BdpA temperature jump measurements to a final temperature of 70 °C which reported a fast ~100 ns phase and a slower ~10 μs phase.[176] For all of the peptides, kinetics of the slow phase of 1632 cm^{-1}, assigned to the solvated helix, are slightly faster than the kinetics of the slow phase of 1648 cm^{-1}, buried helix.

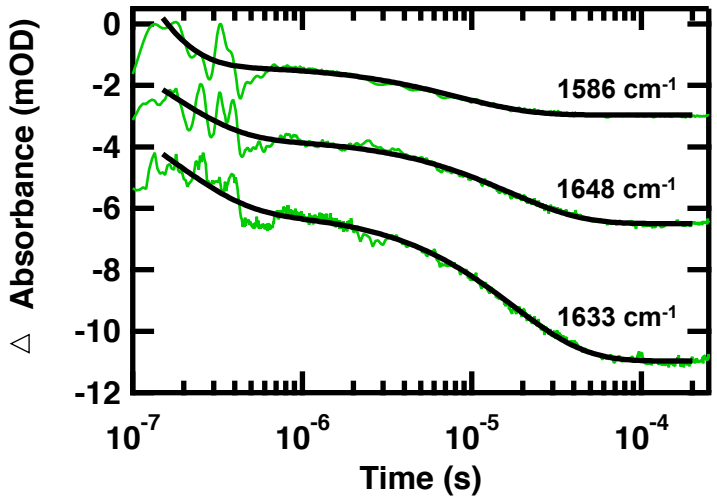


Figure A-7. Representative IR T-jump relaxation kinetics of $^{13}C=^{18}O$ labeled I32M BdpA monitored in the amide I' spectral region at 1648, 1633, and 1586 cm^{-1} following a T-jump from 50 to 60 °C. A double exponential fit is overlaid over each kinetic trace (black solid line). Data are offset for clarity.

Table A-2: Relaxations Kinetics Following a Jump from 50 to 60 °C

	~1648 cm^{-1}		**~1632 cm^{-1}**		**~1560 cm^{-1}**	
	τ_1 (ns)	τ_2 (µs)	τ_1 (ns)	τ_2 (µs)	τ_1 (ns)	τ_2 (µs)
BdpA	115 ± 5	13.2 ± 0.2	185 ± 8	9.4 ± 0.1		
Y15M*	127 ± 5	46.0 ± 0.7	139 ± 4	39.1 ± 0.8	130	25 ± 1
I32M	160 ± 10	16.6 ± 0.3	170 ± 10	16.4 ± 0.3	77 ± 6	8.2 ± 0.3
A47M	53 ± 2	18.7 ± 0.3	110 ± 4	7.4 ± 0.6	120 ± 20	8.0 ± 0.5

*Relaxation kinetics following a jump from 20 to 30 °C. No confidence interval is reported for fast phase of 1560 cm^{-1} because of early timescale artifacts.

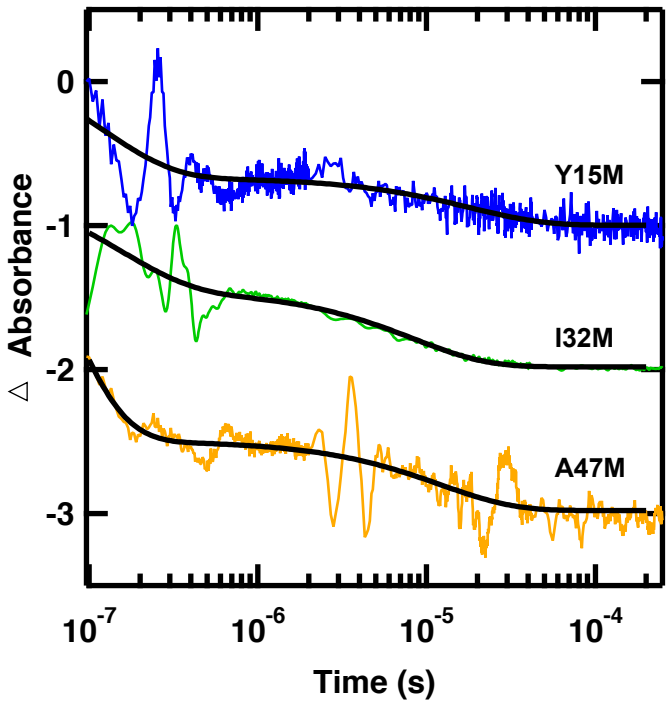


Figure A-8. Representative IR T-jump relaxation kinetics of Y15M BdpA (blue), I32M BdpA (green) and A47M BdpA (yellow) monitored at the $^{13}C=^{18}O$ labeled amide I' spectral position at ~1560 cm^{-1} following a T-jump from 50 to 60 °C (20 to 30 °C for Y15M). A double exponential fit is overlaid on each kinetic trace (black solid line). Data are normalized and offset for clarity.

Temperature jump measurements at 1560 cm^{-1} (Figure A-8) reveal the kinetics at each of the $^{13}C=^{18}O$ labeled positions. Similar to the solvated and buried helix wavelengths, the kinetics of the labeled position fit to a double exponential with a fast ~100 ns phase and a slower ~10 μs phase. There is relatively more noise in this data compared to the data for the unlabeled solvated and buried helix because of the smaller overall amplitude of absorbance at the labeled position (Figure A-4). The kinetics at the labeled position

correspond most closely with the kinetics at the solvated position, supporting the FTIR assignment of these positions to partially solvated helix (Table A-2). Because of early timescale artifacts in the Y15M BdpA data, no confidence interval is reported for the fast phase. Despite the large artifacts present at early times in the Y15M data, the fast phase is clearly present, and its lifetime is of the same order of magnitude as that observed for the other mutants. The kinetics probed at 1648 and 1632 cm^{-1} are dominated by the slow phase but also exhibit a smaller amplitude fast phase. In contrast, the kinetics probed at the labeled frequency 1560 cm^{-1} show nearly equal amplitudes in the fast and slow phases.

The two well-separated relaxation phases make it clear that BdpA folds through an intermediate state as described in Figure A-2. Both the folded and intermediate states are populated at the starting temperature of the T-jump experiments, such that relaxation is observed over both barriers. The fast phase corresponds to relaxation over the first barrier between the U and I states, whereas the slower phase is assigned to the relaxation over the global transition state barrier between I and the fully folded state. The equilibrium CD and FTIR melts are also consistent with a folding intermediate despite reasonable fits to a two-state model, since the T_m and transition width depend on the probe, whether CD or the different IR probe frequencies. A previous infrared study of the WT BdpA studied this phenomenon extensively using T-jumps to multiple temperatures to deconvolve a second broad melt component with a T_m of 58 °C corresponding to the fast kinetics phase.[176] This phase was assigned to fast helical formation preceding formation of long-range tertiary contacts on the microsecond timescale. Differences in the FTIR equilibrium melting temperatures of the solvated and buried helices (Figure A-6) suggest that this

intermediate has a largely intact hydrophobic core with fraying at the ends of the helices. Whereas the transients of the solvated and buried helix have a small intensity in the fast phase, the transients of the $^{13}C=^{18}O$ labeled position have nearly half of their intensity in the fast phase. The $^{13}C=^{18}O$ labels are local probes, whereas the solvated and buried helix positions are more sensitive to global changes. The larger relative intensity in the fast phase when labels are incorporated at positions in the hydrophobic core is indicative of an intermediate in which these positions are helical. The infrared melt probed at the $^{13}C=^{18}O$ labeled position is more broad than the melt at the buried or solvated helix position, supporting our conclusion that the absorbance includes relatively more contribution from the broad melt component. Because of the small overall absorbance intensity at this position it is not possible to deconvolve the two melts, however, the breadth of the infrared melt combined with the observations of two phases from kinetics measurements support the assignment of the fast helical formation phase at the $^{13}C=^{18}O$ labeled position.

While phi-value analysis has been extensively used to identify specific interactions formed in the transition state, there have been few time resolved studies that resolve a fast forming intermediate, which can indirectly report on the transition state structure. The studies that have been conducted rely on incorporation of tryptophan or non-natural fluorescent probes and most do not resolve the nanosecond timescale.[171, 172, 174, 175, 186-188] Many utilize tryptophan as a reporter of global BdpA folding, with probes located only in Helix 1.[172, 175, 187, 188] Previous work from the Dyer group used nanosecond T-jump fluorescence measurements of tryptophan probes at similar positions to the $^{13}C=^{18}O$ methionine mutations in Helix 1, 2 and 3.[174] A fast phase was observed only for the Helix

2 mutant, suggesting that the initial step in BdpA formation is formation of Helix 2. Radford and coworkers also did not observe a fast phase in their nanosecond T-jump fluorescence measurements of a fast folding mutant of BdpA with a probe in Helix 1.[187] However, interpretation of tryptophan fluorescence is difficult and the lack of a fast phase for the Helix 1 and Helix 3 mutants does not preclude formation of some helical structure. If, as in the model proposed by infrared measurements, the fast phase arises from helical formation and the slow phase from tertiary contact formation, changes in the environment necessary for changes in tryptophan fluorescence may not occur until tertiary contact formation.[176] Our results support this model, as we observe a fast phase for infrared probes located within one residue of the tryptophan mutants. We incorporated our infrared probes at locations where phi-value analysis performed by Fersht and co-workers determined transition state structure formation in Helix 1, Helix 2 and Helix 3.[171, 172] Our measurements demonstrate that there is an intermediate with helical structure formed at these positions, providing the first time-resolved evidence of structure formation in Helix 1 and Helix 3 prior to forming the global transition state.

While kinetic rates of BdpA derived from computer simulations agree with experimental results,[195-197] the models do not necessarily capture the details of the folding mechanism correctly. One group of all-atom simulations predict early formation of Helix 3.[180-185] Results from phi-value analysis and fluorescence T-jump conflict with this model; they predict formation of Helix 2 with at most partial formation of contacts in Helix 1 and 3. They agree better with a second group of simulations that predict early formation of Helix 1 and Helix 2.[177-179] The distributed computing approach employed by the Pande group predicts that while Helix 2 is most likely to form first and lead to the native state,

there is an alterative pathway where Helix 3 forms first. [197] Our kinetic measurements provide evidence for this model, as we observe at least partial formation of each of the helices in the intermediate state that precedes formation of the global transition state. Further measurements of BdpA labeled at additional positions in the helices would be necessary to determine fully the degree of helix formation prior to the global transition state. The results demonstrate that $^{13}C=^{18}O$ labeled methionine infrared labels provide a powerful tool for studying protein folding at the single-residue level.

A.6 Conclusion

Using $^{13}C=^{18}O$ labeled methionine infrared labels we are able to resolve folding of individual residues within BdpA. We propose a transition state model where Helix 1, Helix 2, and Helix 3 are all at least partially structured, similar to that proposed by phi-value analysis.[171, 172] This is the first direct observation of fast helix formation in Helix 1 and Helix 3, supporting computational models that predict formation of these helices in the transition state. $^{13}C=^{18}O$ labeled methionine infrared probes are sensitive to structural changes of a specific residue in the protein. This provides complementary information to unlabeled infrared spectroscopy, sensitive to changes in secondary structure of the protein, and fluorescence spectroscopy, which is sensitive to environmental changes around the fluorophore. Recombinant expression in a methionine auxotroph provides an efficient and in the present case minimally perturbing method for incorporating isotopic labels. In principle, any position may be isotopically labeled using recombinant expression, provided it will tolerate a methionine substitution. By utilizing isotope-edited

infrared probes at multiple positions in the protein we have established an experimental methodology to approach an all-atom view of the folding landscape.

A.7 Appendix

1. Temperature dependent FTIR spectra of I32M and A47M BdpA mutants were recorded to confirm that the peptide was folded. Temperature dependent FTIR of I32M and A47M BdpA mutants are reported in Figure AA-1.

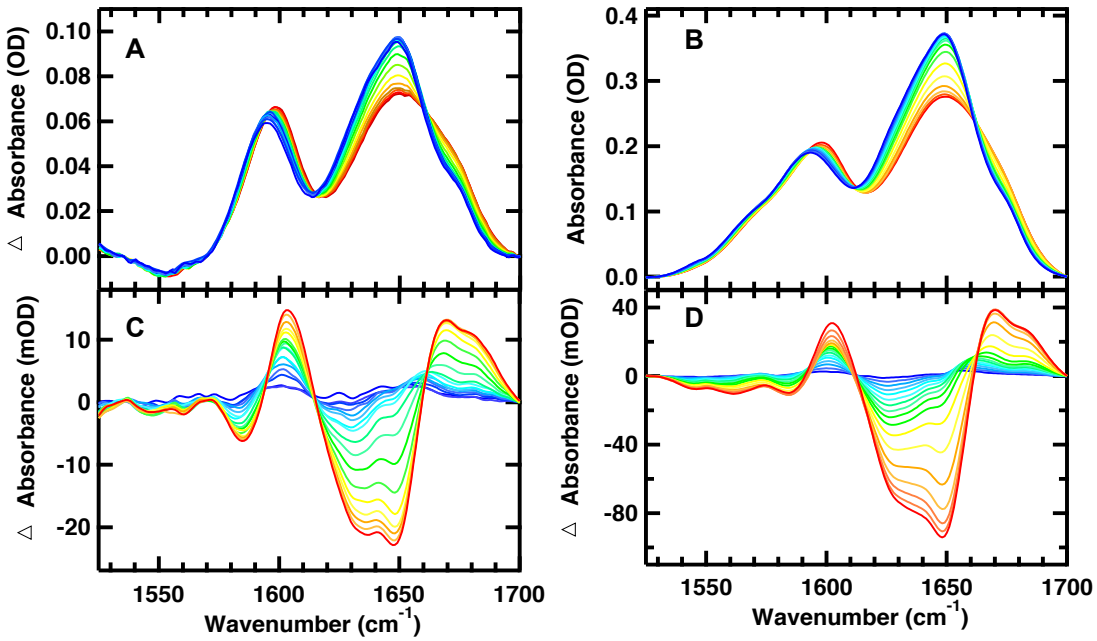


Figure AA-1. Temperature dependent FTIR spectra of 2 mg/ml I32M BdpA (A, C) and 8 mg/ml A47M BdpA (B, D) in 25 mM potassium phosphate and 50 mM NaCl (pH 6.8). (A, B) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 25 to 100 °C in 5 °C intervals. (C, D) Difference spectra obtained by subtracting the spectrum at 25 °C from the spectra at higher temperatures.

2. Protein was treated with $NiCl_2$ to remove EDTA contamination. Supernatent harvested during the removal of the histidine tag from contained some residual EDTA, which interfered in the infrared. The supernatant was treated with a 1:1 ratio of $NiCl_2$ to EDTA in order to remove any EDTA in the solution. The solution was dialyzed with water overnight to remove the NiEDTA product and buffer solution. Dialyzed sample was lyophilized and exchanged in D_2O. The Y15M mutant pre and post cleaning is reported in Figure AA-2. The IR peaks are more easily distinguished in the second derivative spectra (Figure AA-3). IR peaks assigned to solvated (1632 cm^{-1}) and buried (1648 cm^{-1}) helix and the labeled position (1560 cm^{-1}) can be identified in both the pre and post cleaned Y15M sample. The peak arising from residual EDTA is centered ~20 cm^{-1} away from the labeled position, so it does not interfere in the region of interest. A normalized melt of the labeled position pre and post purification reveals that the melting temperature and cooperativity of folding is unchanged (Figure AA-3 B).

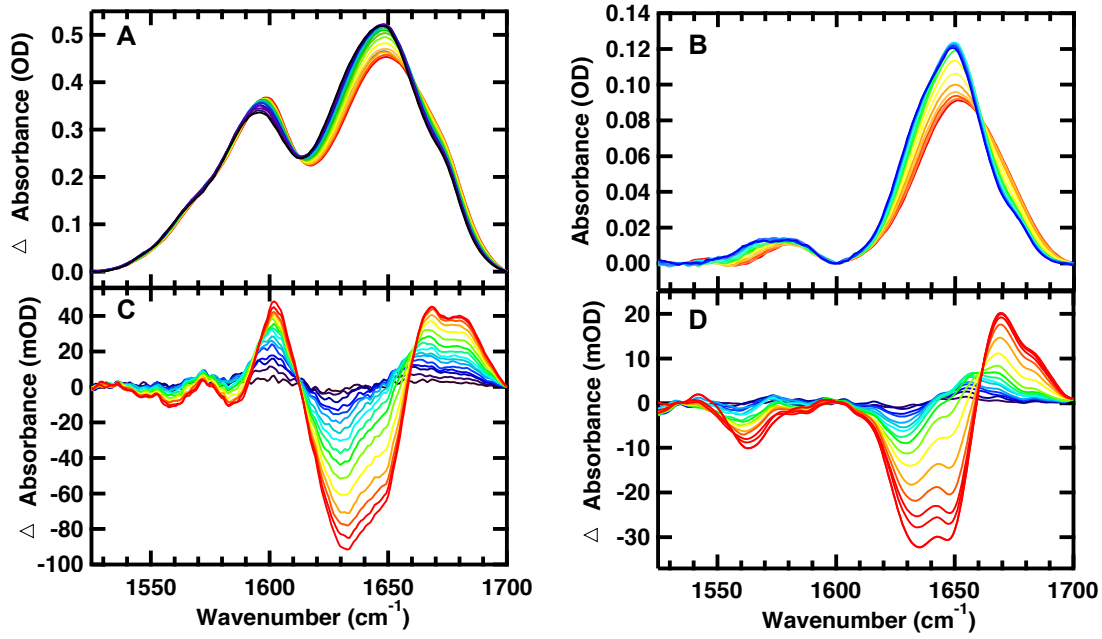


Figure AA-2. Temperature dependent FTIR spectra of 8 mg/ml Y15M BdpA with EDTA contamination (A, C) and 3 mg/ml Y15M BdpA without EDTA contamination (B, D) in 25 mM potassium phosphate and 50 mM NaCl (pH 6.8). (A, B) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 25 to 100 °C in 5 °C intervals. (C, D) Difference spectra obtained by subtracting the spectrum at 25 °C from the spectra at higher temperatures.

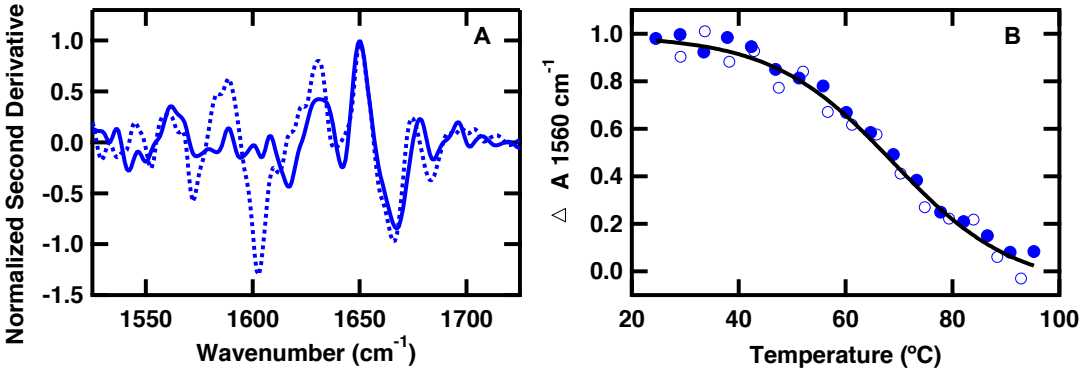


Figure AA-3. Data are normalized at the maximum for clarity. (A) Second derivative of the FTIR difference spectrum (100-25 °C) of Y15M BdpA pre (dashed line) and post (solid line) removal of EDTA. (B) FTIR melt curves of Y15M pre (open circles) and post (closed circles) removal of EDTA obtained by plotting the change in IR difference spectra at 1560 cm^{-1} versus temperature.

3. Temperature dependent FTIR spectra of A47M and $^{13}C{=}^{18}O$ were collected to confirm the position of the labeled peak. Temperature dependent FTIR of His-Tagged A47M and His-Tagged $^{13}C{=}^{18}O$ A47M BdpA mutants are reported in Figure AA-4. A second derivative of the lowest temperature absorbance data (Figure AA-5) reveals that there is good agreement in the folded structure of the His-Tagged A47M, wildtype and $^{13}C{=}^{18}O$ His-Tagged A47M. There are no differences in the side chain contributions from the folded His-Tagged A47M mutant compared to the wildtype mutant.

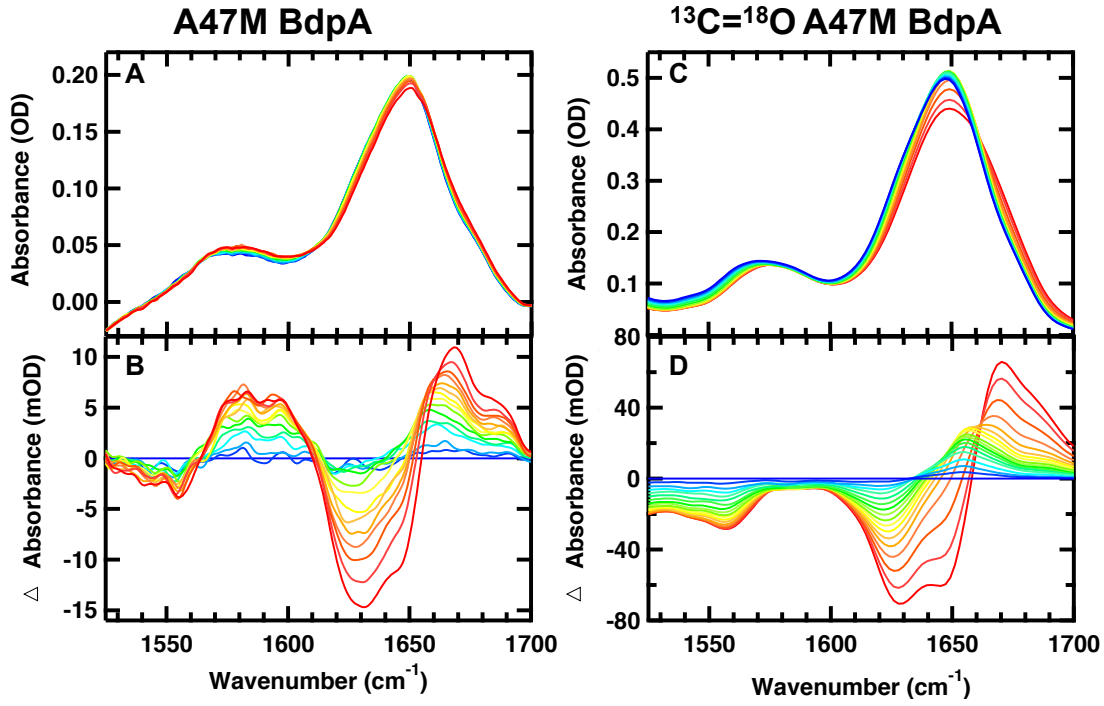


Figure AA-4. Temperature dependent FTIR spectra of His-Tagged A47M BdpA (A,C) and His-Tagged $^{13}C=^{18}O$ A47M BdpA (B, D) in 25 mM potassium phosphate and 50 mM NaCl (pH 6.8). (A, B) Absorbance spectra in Amide I' region; the temperatures of the individual traces varies from 25 to 100 °C in 5 °C intervals. (C, D) Difference spectra obtained by subtracting the spectrum at 25 °C from the spectra at higher temperatures

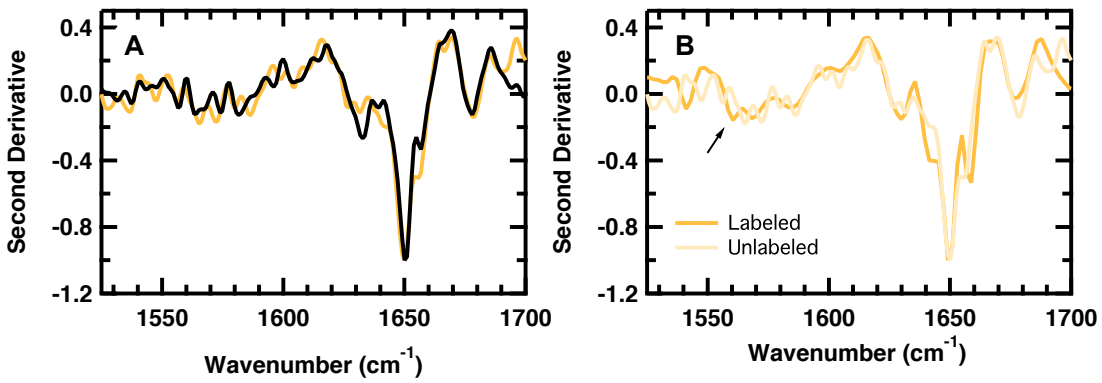


Figure AA-5. Second derivative of the FTIR spectra collected at 25 °C. (A) Comparison of WT (black) and His-Tagged A47M (Yellow). (B) Comparison of His-Tagged $^{13}C=^{18}O$ A47M (Dark Yellow) and His-Tagged A47M. Arrow points to 1560 cm^{-1} peak.

4. IR T-jump of the wildtype, Y15M and A47M BdpA were collected to determine the dynamics of the systems. Wavelength dependent IR T-jump measurements were collected at ~1648, 1632 and 1560 cm^{-1}. Representative IR T-jump relaxation kinetics of wildtype, Y15M and A47M BdpA are reported in Figure AA-6, AA-7, AA-8.

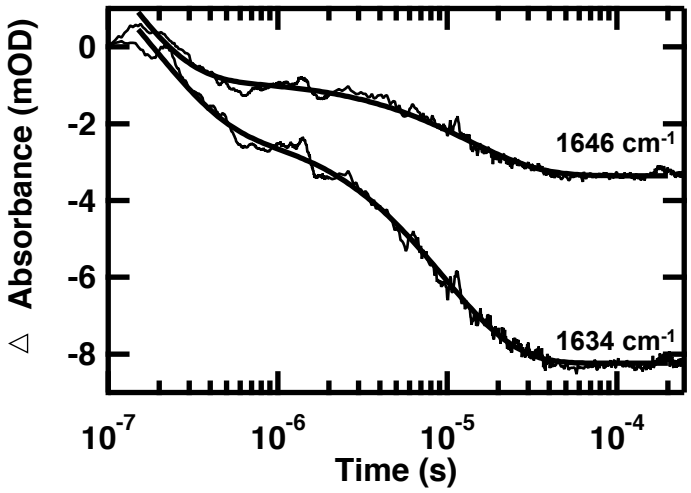


Figure AA-6. Representative IR T-jump relaxation kinetics of wildtype BdpA monitored in the amide I' spectral region at 1646 and 1634 cm^{-1} following a T-jump from 50 to 60 °C. A double exponential fit is overlaid over each kinetic trace (black solid line).

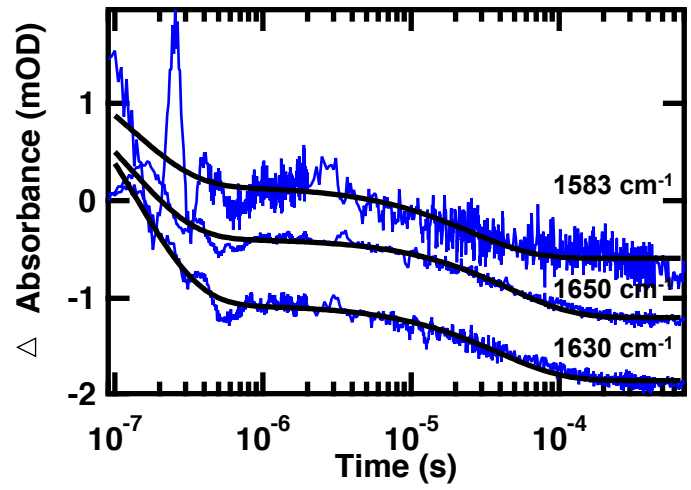


Figure AA-7. Representative IR T-jump relaxation kinetics of Y15M BdpA monitored in the amide I' spectral region at 1650, 1630 and 1583 cm^{-1} following a T-jump from 20 to 30 °C. A double exponential fit is overlaid over each kinetic trace (black solid line).

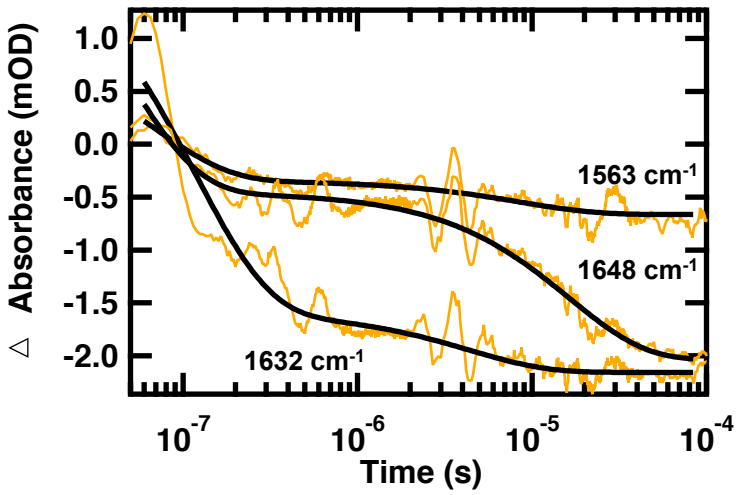


Figure AA-8. Representative IR T-jump relaxation kinetics of A47M BdpA monitored in the amide I' spectral region at 1648, 1632 and 1563 cm^{-1} following a T-jump from 50 to 60 °C. A double exponential fit is overlaid over each kinetic trace (black solid line).

REFERENCES

1. Pennisi, E. (2012) Genomics. Encode project writes eulogy for junk DNA, *Science 337*, 1159, 1161.

2. Callender, R. H., Dyer, R. B., Gilmanshin, R., and Woodruff, W. H. (1998) Fast events in protein folding: The time evolution of primary processes, *Annual review of physical chemistry 49*, 173-202.

3. Dobson, C. M. (2002) Getting out of shape-protein misfolding diseases *Nature 418*, 729-730.

4. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem, *Annual review of biophysics 37*, 289-316.

5. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-230.

6. Finkelstein, A. V. (1997) Protein structure: What is it possible to predict now?, *Curr Opin Struct Biol 7*, 60-71.

7. Levintha.C. (1968) Are there pathways for protein folding, *Journal De Chimie Physique Et De Physico-Chimie Biologique 65*, 44-&.

8. Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins, *Proc Natl Acad Sci U S A 70*, 697-701.

9. Karplus, M., and Weaver, D. L. (1976) Protein-folding dynamics, *Nature 260*, 404-406.

10. Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on, *Science 338*, 1042-1046.

11. Baldwin, R. L. (1989) How does protein folding get started?, *Trends in biochemical sciences 14*, 291-294.

12. Dill, K. A. (1985) Theory for the folding and stability of globular proteins, *Biochemistry-Us 24*, 1501-1509.

13. Harrison, S. C., and Durbin, R. (1985) Is there a single pathway for the folding of a polypeptide chain?, *Proc Natl Acad Sci U S A 82*, 4028-4030.

14. Bryngelson, J. D., and Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding, *Proc Natl Acad Sci U S A 84*, 7524-7528.

15. Dill, K. A. (1999) Polymer principles and protein folding, *Protein Sci 8*, 1166-1180.

16. Dill, K. A., Alonso, D. O., and Hutchinson, K. (1989) Thermal stabilities of globular proteins, *Biochemistry-Us 28*, 5439-5449.

17. Dill, K. A., and Chan, H. S. (1997) From levinthal to pathways to funnels, *Nat Struct Biol 4*, 10-19.

18. Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship, *Proc Natl Acad Sci U S A 89*, 8721-8725.

19. Dyer, R. B. (2007) Ultrafast and downhill protein folding, *Curr Opin Struc Biol 17*, 38-47.

20. Ode, H., Nakashima, M., Kitamura, S., Sugiura, W., and Sato, H. (2012) Molecular dynamics simulation in virus research, *Frontiers in microbiology 3*, 258.

21. Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold, *Science 334*, 517-520.

22. Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., Rhee, Y. M., Shirts, M. R., Snow, C. D., Sorin, E. J., and Zagrovic, B. (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing, *Biopolymers 68*, 91-109.

23. Snow, C. D., Sorin, E. J., Rhee, Y. M., and Pande, V. S. (2005) How well can simulation predict protein folding kinetics and thermodynamics?, *Annu. Rev. Biophys. Biomolec. Struct. 34*, 43-69.

24. Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding 'speed limit', *Curr Opin Struc Biol 14*, 76-88.

25. Chan, C. K., Hofrichter, J., and Eaton, W. A. (1996) Optical triggers of protein folding, *Science 274*, 628-629.

26. Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W. A. (1993) Fast events in protein folding initiated by nanosecond laser photolysis, *Proc Natl Acad Sci U S A 90*, 11860-11864.

27. Knight, J. B., Vishwanath, A., Brody, J. P., and Austin, R. H. (1998) Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds, *Physical review letters 80*, 3863-3866.

28. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing, *Proc Natl Acad Sci U S A 94*, 1779-1784.

29. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J Mol Biol 277*, 985-994.

30. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1996) Fast events in protein folding: Helix melting and formation in a small peptide, *Biochemistry-Us 35*, 691-697.

31. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) Laser temperature jump study of the helix<==>coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model, *Biochemistry-Us 36*, 9200-9210.

32. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding, *Proc Natl Acad Sci U S A 93*, 11615-11617.

33. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis, *Proteins 21*, 167-195.

34. Yang, W. Y., and Gruebele, M. (2004) Folding lambda-repressor at its speed limit, *Biophys J 87*, 596-608.

35. Kortemme, T., Ramirez-Alvarado, M., and Serrano, L. (1998) Design of a 20-amino acid, three-stranded beta-sheet protein, *Science 281*, 253-256.

36. Ramirez-Alvarado, M., Kortemme, T., Blanco, F. J., and Serrano, L. (1999) Beta-hairpin and beta-sheet formation in designed linear peptides, *Bioorganic & medicinal chemistry 7*, 93-103.

37. Blanco, F., Ramirez-Alvarado, M., and Serrano, L. (1998) Formation and stability of beta-hairpin structures in polypeptides, *Curr Opin Struct Biol 8*, 107-111.

38. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of beta-hairpin formation, *Nature 390*, 196-199.

39. Luo, Z. L., Ding, J. D., and Zhou, Y. Q. (2008) Folding mechanisms of individual beta-hairpins in a go model of pin1 ww domain by all-atom molecular dynamics simulations, *Journal of Chemical Physics 128*.

40. Barth, A., and Zscherp, C. (2002) What vibrations tell us about proteins, *Q Rev Biophys 35*, 369-430.

41. Davis, C. M., Xiao, S. F., Raleigh, D. P., and Dyer, R. B. (2012) Raising the speed limit for beta-hairpin formation, *J Am Chem Soc 134*, 14476-14482.

42. Eaton, W. A., Munoz, V., Thompson, P. A., Henry, E. R., and Hofrichter, J. (1998) Kinetics and dynamics of loops, alpha-helices, beta-hairpins, and fast-folding proteins, *Accounts Chem Res 31*, 745-753.

43. Werner, J. H., Dyer, R. B., Fesinmeyer, R. M., and Andersen, N. H. (2002) Dynamics of the primary processes of protein folding: Helix nucleation, *J. Phys. Chem. B 106*, 487-494.

44. Andersen, N. H., Fesinmeyer, R. M., Peterson, E. S., and Dyer, R. B. (2005) Studies of helix fraying and solvation using c-13 ' isotopomers, *Protein Sci. 14*, 2324-2332.

45. Dyer, R. B., and Causgrove, T. P. (2006) Nonequilibrium protein folding dynamics: Laser-induced ph-jump studies of the-helix-coil transition, *Chem Phys 323*, 2-10.

46. Dinner, A. R., Lazaridis, T., and Karplus, M. (1999) Understanding beta-hairpin formation, *Proc. Natl. Acad. Sci. U. S. A. 96*, 9068-9073.

47. Klimov, D. K., and Thirumalai, D. (2000) Mechanisms and kinetics of beta-hairpin formation, *P Natl Acad Sci USA 97*, 2544-2549.

48. Pande, V. S., and Rokhsar, D. S. (1999) Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein g, *P Natl Acad Sci USA 96*, 9062-9067.

49. Blanco, F. J., Rivas, G., and Serrano, L. (1994) A short linear peptide that folds into a native stable beta-hairpin in aqueous solution, *Nat. Struct. Biol. 1*, 584-590.

50. Maness, S. J., Franzen, S., Gibbs, A. C., Causgrove, T. P., and Dyer, R. B. (2003) Nanosecond temperature jump relaxation dynamics of cyclic beta-hairpin peptides, *Biophys J 84*, 3874-3882.

51. Dyer, R. B., Maness, S. J., Peterson, E. S., Franzen, S., Fesinmeyer, R. M., and Andersen, N. H. (2004) The mechanism of beta-hairpin formation, *Biochemistry 43*, 11560-11566.

52. Yang, W. Y., and Gruebele, M. (2004) Detection-dependent kinetics as a probe of folding landscape microstructure, *J Am Chem Soc 126*, 7758-7759.

53. Xu, Y., Oyola, R., and Gai, F. (2003) Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin, *Journal of the American Chemical Society 125*, 15388-15394.

54. Honda, S., Akiba, T., Kato, Y. S., Sawada, Y., Sekijima, M., Ishimura, M., Ooishi, A., Watanabe, H., Odahara, T., and Harata, K. (2008) Crystal structure of a ten-amino acid protein, *J Am Chem Soc 130*, 15327-15331.

55. Honda, S., Kobayashi, N., and Munekata, E. (2000) Thermodynamics of a beta-hairpin structure: Evidence for cooperative formation of folding nucleus, *J Mol Biol 295*, 269-278.

56. Honda, S., Yamasaki, K., Sawada, Y., and Morii, H. (2004) 10 residue folded peptide designed by segment statistics, *Structure 12*, 1507-1518.

57. Kitao, A. (2011) Transform and relax sampling for highly anisotropic systems: Application to protein domain motion and folding (vol 135, 045101, 2011), *J Chem Phys 135*.

58. Harada, R., and Kitao, A. (2011) Exploring the folding free energy landscape of a beta-hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method, *J. Phys. Chem. B 115*, 8806-8812.

59. Suenaga, A., Narumi, T., Futatsugi, N., Yanai, R., Ohno, Y., Okimoto, N., and Taiji, M. (2007) Folding dynamics of 10-residue beta-hairpin peptide chignolin, *Chem-Asian J 2*, 591-598.

60. Munoz, V., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1998) A statistical mechanical model for beta-hairpin kinetics, *P Natl Acad Sci USA 95*, 5872-5879.

61. Gulotta, M., Rogatsky, E., Callender, R. H., and Dyer, R. B. (2003) Primary folding dynamics of sperm whale apomyoglobin: Core formation, *Biophys. J. 84*, 1909-1918.

62. Yang, W. J., Griffiths, P. R., Byler, D. M., and Susi, H. (1985) Protein conformation by infrared-spectroscopy - resolution enhancement by fourier self-deconvolution, *Appl Spectrosc 39*, 282-287.

63. Arrondo, J. L. R., Blanco, F. J., Serrano, L., and Goni, F. M. (1996) Infrared evidence of a beta-hairpin peptide structure in solution, *Febs Lett 384*, 35-37.

64. Susi, H., and Byler, D. M. (1986) Resolution-enhanced fourier-transform infrared-spectroscopy of enzymes, *Method Enzymol 130*, 290-311.

65. Dyer, R. B., Gai, F., and Woodruff, W. H. (1998) Infrared studies of fast events in protein folding, *Accounts Chem Res 31*, 709-716.

66. Hilario, J., Kubelka, J., and Keiderling, T. A. (2003) Optical spectroscopic investigations of model beta-sheet hairpins in aqueous solution, *J Am Chem Soc 125*, 7562-7574.

67. Wang, L., Middleton, C. T., Zanni, M. T., and Skinner, J. L. (2011) Development and validation of transferable amide i vibrational frequency maps for peptides, *J. Phys. Chem. B 115*, 3713-3724.

68. Brewer, S. H., Vu, D. M., Tang, Y., Li, Y., Franzen, S., Raleigh, D. P., and Dyer, R. B. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain, *Proc. Natl. Acad. Sci. U. S. A. 102*, 16662-16667.

69. Hagen, S. J., Hofrichter, J., and Eaton, W. A. (1997) Rate of intrachain diffusion of unfolded cytochrome c, *J. Phys. Chem. B 101*, 2352-2365.

70. Davis, C. M., and Dyer, R. B. (2014) Ww domain folding complexity revealed by infrared spectroscopy, *Biochemistry-Us 53*, 5476-5484.

71. Culik, R. M., Serrano, A. L., Bunagan, M. R., and Gai, F. (2011) Achieving secondary structural resolution in kinetic measurements of protein folding: A case study of the folding mechanism of trp-cage, *Angew Chem Int Edit 50*, 10884-10887.

72. Vu, D. M., Brewer, S. H., and Dyer, R. B. (2012) Early turn formation and chain collapse drive fast folding of the major cold shock protein cspa of escherichia coli, *Biochemistry-Us 51*, 9104-9111.

73. Macias, M. J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) Structure of the ww domain of a kinase-associated protein complexed with a proline-rich peptide, *Nature 382*, 646-649.

74. Macias, M. J., Gervais, V., Civera, C., and Oschkinat, H. (2000) Structural analysis of ww domains and design of a ww prototype, *Nat Struct Biol 7*, 375-379.

75. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group ivww domains, *Nat Struct Biol 7*, 639-643.

76. Wiesner, S., Stier, G., Sattler, M., and Macias, M. J. (2002) Solution structure and ligand recognition of the ww domain pair of the yeast splicing factor prp40, *J Mol Biol 324*, 807-822.

77. Piana, S., Sarkar, K., Lindorff-Larsen, K., Guo, M. H., Gruebele, M., and Shaw, D. E. (2011) Computational design and experimental testing of the fastest-folding beta-sheet protein, *J Mol Biol 405*, 43-48.

78. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W., and Gruebele, M. (2001) The folding mechanism of a beta-sheet: The ww domain, *J Mol Biol 311*, 373-393.

79. Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H., and Fersht, A. (2001) Ultrafast folding of ww domains without structured aromatic clusters in the denatured state, *P Natl Acad Sci USA 98*, 13002-13007.

80. Karanicolas, J., and Brooks, C. L., 3rd. (2003) The structural basis for biphasic kinetics in the folding of the ww domain from a formin-binding protein: Lessons for protein design?, *Proc Natl Acad Sci U S A 100*, 3954-3959.

81. Karanicolas, J., and Brooks, C. L. (2004) Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a ww domain, *P Natl Acad Sci USA 101*, 3432-3437.

82. Mu, Y. G., Nordenskiold, L., and Tam, J. P. (2006) Folding, misfolding, and amyloid protofibril formation of ww domain fbp28, *Biophys J 90*, 3983-3992.

83. Noe, F., Schutte, C., Vanden-Eijnden, E., Reich, L., and Weikl, T. R. (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proc Natl Acad Sci U S A 106*, 19011-19016.

84. Nguyen, H., Jager, M., Moretto, A., Gruebele, M., and Kelly, J. W. (2003) Tuning the free-energy landscape of a ww domain by temperature, mutation, and truncation, *P Natl Acad Sci USA 100*, 3948-3953.

85. Ferguson, N., Berriman, J., Petrovich, M., Sharpe, T. D., Finch, J. T., and Fersht, A. R. (2003) Rapid amyloid fiber formation from the fast-folding ww domain fbp28, *P Natl Acad Sci USA 100*, 9814-9819.

86. Petrovich, M., Jonsson, A. L., Ferguson, N., Daggett, V., and Fersht, A. R. (2006) Phi-analysis at the experimental limits: Mechanism of beta-hairpin formation, *J Mol Biol 360*, 865-881.

87. Liu, F., Du, D. G., Fuller, A. A., Davoren, J. E., Wipf, P., Kelly, J. W., and Gruebele, M. (2008) An experimental survey of the transition between two-state and downhill protein folding scenarios, *P Natl Acad Sci USA 105*, 2369-2374.

88. Davis, C. M., and Dyer, R. B. (2013) Dynamics of an ultrafast folding subdomain in the context of a larger protein fold, *J Am Chem Soc 135*, 19260-19267.

89. Chung, H. S., McHale, K., Louis, J. M., and Eaton, W. A. (2012) Single-molecule fluorescence experiments determine protein folding transition path times, *Science 335*, 981-984.

90. Andrushchenko, V. V., Vogel, H. J., and Prenner, E. J. (2007) Optimization of the hydrochloric acid concentration used for trifluoroacetate removal from synthetic peptides, *J Pept Sci 13*, 37-43.

91. Manning, M. C., and Woody, R. W. (1987) Theoretical determination of the cd of proteins containing closely packed antiparallel beta-sheets, *Biopolymers 26*, 1731-1752.

92. Koepf, E. K., Petrassi, H. M., Sudol, M., and Kelly, J. W. (1999) Ww: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and gdnhcl and a disordered thermal unfolded state, *Protein Sci 8*, 841-853.

93. Koepf, E. K., Petrassi, H. M., Ratnaswamy, G., Huff, M. E., Sudol, M., and Kelly, J. W. (1999) Characterization of the structure and function of w -> fww domain variants: Identification of a natively unfolded protein that folds upon ligand binding, *Biochemistry-Us 38*, 14338-14351.

94. Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M., and Serrano, L. (1994) Characterization of the interaction of natural proline-rich peptides with 5 different sh3 domains, *Biochemistry-Us 33*, 10925-10933.

95. Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I., and Ladenstein, R. (1998) Thermal unfolding of small proteins with sh3 domain folding pattern, *Proteins 31*, 309-319.

96. Reid, K. L., Rodriguez, H. M., Hillier, B. J., and Gregoret, L. M. (1998) Stability and folding properties of a model beta-sheet protein, escherichia coli cspa, *Protein Sci 7*, 470-479.

97. Kraemer-Pecore, C. M., Lecomte, J. T. J., and Desjarlais, J. R. (2003) A de novo redesign of the ww domain, *Protein Sci 12*, 2194-2205.

98. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free-energy changes determined by the linear extrapolation method .1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry-Us 27*, 8063-8068.

99. Tremmel, S., Beyermann, M., Oschkinat, H., Bienert, M., Naumann, D., and Fabian, H. (2005) C-13-labeled tyrosine residues as local ir probes for monitoring conformational changes in peptides and proteins, *Angew Chem Int Edit 44*, 4631-4635.

100. Wang, T., Xu, Y., Du, D. G., and Gai, F. (2004) Determining beta-sheet stability by fourier transform infrared difference spectra, *Biopolymers 75*, 163-172.

101. Chan, D. C., Bedford, M. T., and Leder, P. (1996) Formin binding proteins bear wwp/ww domains that bind proline-rich peptides and functionally resemble sh3 domains, *Embo J 15*, 1045-1054.

102. Kubelka, J., and Keiderling, T. A. (2001) Differentiation of beta-sheet-forming structures: Ab initio-based simulations of ir absorption and vibrational cd for model peptide and protein beta-sheets, *J Am Chem Soc 123*, 12048-12058.

103. Deechongkit, S., Nguyen, H., Powers, E. T., Dawson, P. E., Gruebele, M., and Kelly, J. W. (2004) Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics, *Nature 430*, 101-105.

104. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding, *Proteins 78*, 2725-2737.

105. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein, *Proc Natl Acad Sci U S A 106*, 12289-12294.

106. Levitt, M., Gerstein, M., Huang, E., Subbiah, S., and Tsai, J. (1997) Protein folding: The endgame, *Annual review of biochemistry 66*, 549-579.

107. Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules .1. Why denaturation of globular protein is a 1st-order phase-transition, *Biopolymers 28*, 1667-1680.

108. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct nmr evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease-a, *Nature 375*, 513-515.

109. Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow nmr-spectroscopy - real-time unfolding studies of 6-f-19-tryptophan-labeled escherichia-coli dihydrofolate-reductase, *P Natl Acad Sci USA 92*, 9318-9322.

110. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain, *Proc Natl Acad Sci U S A 107*, 4955-4960.

111. Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000) Fast kinetics and mechanisms in protein folding, *Annu Rev Biophys Biomol Struct 29*, 327-359.

112. Takada, S. (2001) Protein folding simulation with solvent-induced force field: Folding pathway ensemble of three-helix-bundle proteins, *Proteins-Structure Function and Genetics 42*, 85-98.

113. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds, *Nature 421*, 863-867.

114. Bai, Y. W., Karimi, A., Dyson, H. J., and Wright, P. E. (1997) Absence of a stable intermediate on the folding pathway of protein a, *Protein Sci. 6*, 1449-1457.

115. Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002) Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing (vol 323, pg 927, 2002), *J. Mol. Biol. 324*, 1051-1051.

116. Ferguson, N., Pires, J. R., Toepert, F., Johnson, C. M., Pan, Y. P., Volkmer-Engert, R., Schneider-Mergener, J., Daggett, V., Oschkinat, H., and Fersht, A. (2001) Using flexible loop mimetics to extend phi-value analysis to secondary structure interactions, *Proceedings of the National Academy of Sciences of the United States of America 98*, 13008-13013.

117. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M., and Kelly, J. W. (2006) Structure-function-folding relationship in a ww domain, *Proc Natl Acad Sci U S A 103*, 10648-10653.

118. S, A. B., Skrbic, T., Covino, R., and Faccioli, P. (2012) Dominant folding pathways of a ww domain, *Proc Natl Acad Sci U S A 109*, 2330-2335.

119. Xu, D., and Zhang, Y. (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field, *Proteins-Structure Function and Bioinformatics 80*, 1715-1735.

120. Zhang, Y., and Skolnick, J. (2005) Tm-align: A protein structure alignment algorithm based on the tm-score, *Nucleic Acids Res. 33*, 2302-2309.

121. Lindorff-Larsen, K., Vendruscolo, M., Paci, E., and Dobson, C. M. (2004) Transition states for protein folding have native topologies despite high structural variability, *Nat Struct Mol Biol 11*, 443-449.

122. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) Long-range order in the src sh3 folding transition state, *P Natl Acad Sci USA 97*, 7084-7089.

123. McCallister, E. L., Alm, E., and Baker, D. (2000) Critical role of beta-hairpin formation in protein g folding, *Nature Structural Biology 7*, 669-673.

124. Nauli, S., Kuhlman, B., and Baker, D. (2001) Computer-based redesign of a protein folding pathway, *Nature Structural Biology 8*, 602-605.

125. Oliveberg, M., and Fersht, A. R. (1996) A new approach to the study of transient protein conformations: The formation of a semiburied salt link in the folding pathway of barnase, *Biochemistry-Us 35*, 6795-6805.

126. Stoycheva, A. D., Onuchic, J. N., and Brooks, C. L. (2003) Effect of gatekeepers on the early folding kinetics of a model beta-barrel protein, *Journal of Chemical Physics 119*, 5722-5729.

127. Waldburger, C. D., Jonsson, T., and Sauer, R. T. (1996) Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure, *P Natl Acad Sci USA 93*, 2629-2634.

128. Lane, T. J., Bowman, G. R., Beauchamp, K., Voelz, V. A., and Pande, V. S. (2011) Markov state model reveals folding and functional dynamics in ultra-long md trajectories, *J Am Chem Soc 133*, 18413-18419.

129. Jaeger, M., Nguyen, H., Dendle, M., Gruebele, M., and Kelly, J. W. (2007) Influence of hpin1 wwn-terminal domain boundaries on function, protein stability, and folding, *Protein Sci 16*, 1495-1501.

130. Jager, M., Dendle, M., and Kelly, J. W. (2009) Sequence determinants of thermodynamic stability in a ww domain--an all-beta-sheet protein, *Protein Sci 18*, 1806-1813.

131. Hendler, R. W., and Shrager, R. I. (1994) Deconvolutions based on singular value decomposition and the pseudoinverse: A guide for beginners, *Journal of biochemical and biophysical methods 28*, 1-33.

132. Brewer, S. H., Tang, Y., Vu, D. M., Gnanakaran, S., Raleigh, D. P., and Dyer, R. B. (2012) Temperature dependence of water interactions with the amide carbonyls of alpha-helices, *Biochemistry-Us 51*, 5293-5299.

133. Jager, M., Dendle, M., Fuller, A. A., and Kelly, J. W. (2007) A cross-strand trp trp pair stabilizes the hpin1 ww domain at the expense of function, *Protein Sci 16*, 2306-2313.

134. Nicholson, E. M., and Scholtz, J. M. (1996) Conformational stability of the escherichia coli hpr protein: Test of the linear extrapolation method and a thermodynamic characterization of cold denaturation, *Biochemistry-Us 35*, 11369-11378.

135. Ranganathan, R., Lu, K. P., Hunter, T., and Noel, J. P. (1997) Structural and functional analysis of the mitotic rotamase pin1 suggests substrate recognition is phosphorylation dependent, *Cell 89*, 875-886.

136. Deacon, G. B., and Phillips, R. J. (1980) Relationships between the carbon-oxygen stretching frequencies of carboxylato complexes and the type of carboxylate coordination, *Coordin Chem Rev 33*, 227-250.

137. Urry, D. W., Peng, S. Q., Parker, T. M., Gowda, D. C., and Harris, R. D. (1993) Relative significance of electrostatic-induced and hydrophobic-induced pk(a) shifts in a model protein - the aspartic-acid residue, *Angewandte Chemie-International Edition in English 32*, 1440-1442.

138. Langsetmo, K., Fuchs, J. A., and Woodward, C. (1991) The conserved, buried aspartic-acid in oxidized escherichia-coli thioredoxin has a pka of 7.5 - its titration produces a related shift in global stability, *Biochemistry-Us 30*, 7603-7609.

139. Langsetmo, K., Fuchs, J. A., Woodward, C., and Sharp, K. A. (1991) Linkage of thioredoxin stability to titration of ionizable groups with perturbed pka, *Biochemistry-Us 30*, 7609-7614.

140. Auzat, I., and Garel, J. R. (1992) Ph-dependence of the reverse reaction catalyzed by phosphofructokinase-i from escherichia-coli - implications for the role of asp-127, *Protein Sci 1*, 254-258.

141. Zou, Y., Li, Y. Y., Hao, W. Y., Hu, X. Q., and Ma, G. (2013) Parallel beta-sheet fibril and antiparallel beta-sheet oligomer: New insights into amyloid formation of

hen egg white lysozyme under heat and acidic condition from ftir spectroscopy, *J Phys Chem B 117*, 4003-4013.

142. Kellogg, E. H., Lange, O. F., and Baker, D. (2012) Evaluation and optimization of discrete state models of protein folding, *J Phys Chem B 116*, 11405-11413.

143. Krivov, S. V. (2011) The free energy landscape analysis of protein (fip35) folding dynamics, *J Phys Chem B 115*, 12315-12324.

144. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010) Atomic-level characterization of the structural dynamics of proteins, *Science 330*, 341-346.

145. Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: Globocan 2008, *Int J Cancer 127*, 2893-2917.

146. Harris, F., Dennison, S. R., Singh, J., and Phoenix, D. A. (2013) On the selectivity and efficacy of defense peptides with respect to cancer cells, *Medicinal research reviews 33*, 190-234.

147. Al-Benna, S., Shai, Y., Jacobsen, F., and Steinstraesser, L. (2011) Oncolytic activities of host defense peptides, *Int J Mol Sci 12*, 8027-8051.

148. Hoskin, D. W., and Ramamoorthy, A. (2008) Studies on anticancer activities of antimicrobial peptides, *Bba-Biomembranes 1778*, 357-375.

149. Schweizer, F. (2009) Cationic amphiphilic peptides with cancer-selective toxicity, *Eur J Pharmacol 625*, 190-194.

150. Leuschner, C., and Hansel, W. (2004) Membrane disrupting lytic peptides for cancer treatments, *Curr Pharm Design 10*, 2299-2310.

151. Papo, N., and Shai, Y. (2005) Host defense peptides as new weapons in cancer treatment, *Cell Mol Life Sci 62*, 784-790.

152. Papo, N., and Shai, Y. (2003) New lytic peptides based on the d,l-amphipathic helix motif preferentially kill tumor cells compared to normal cells, *Biochemistry-Us 42*, 9346-9354.

153. Mader, J. S., and Hoskin, D. W. (2006) Cationic antimicrobial peptides as novel cytotoxic agents for cancer treatment, *Expert Opin Inv Drug 15*, 933-946.

154. Sinthuvanich, C., Veiga, A. S., Gupta, K., Gaspar, D., Blumenthal, R., and Schneider, J. P. (2012) Anticancer beta-hairpin peptides: Membrane-induced folding triggers activity, *J Am Chem Soc 134*, 6210-6217.

155. Venyaminov, S., Baikalov, I. A., Shen, Z. M., Wu, C. S., and Yang, J. T. (1993) Circular dichroic analysis of denatured proteins: Inclusion of denatured proteins in the reference set, *Analytical biochemistry 214*, 17-24.

156. Silvius, J. R. (1982) Thermotropic phase transitions of pure lipids in model membranes and their modifications by membrane proteins, In *Lipid-protein interactions*, John Wiley & Sons, New York.

157. Christiaens, B., Symoens, S., Vanderheyden, S., Engelborghs, Y., Joliot, A., Prochiantz, A., Vandekerckhove, J., Rosseneu, M., and Vanloo, B. (2002) Tryptophan fluorescence study of the interaction of penetratin peptides with model membranes, *Eur J Biochem 269*, 2918-2926.

158. Beechem, J. M., and Brand, L. (1985) Time-resolved fluorescence of proteins, *Annual review of biochemistry 54*, 43-71.

159. Weber, G. (1960) Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds, *Biochem J 75*, 335-345.

160. Eftink, M. R. (1991) Fluorescence techniques for studying protein-structure, *Method Biochem Anal 35*, 127-205.

161. Kirby, E. P., and Steiner, R. F. (1970) Influence of solvent and temperature upon the fluorescence of indole derivatives, *The Journal of Physical Chemistry 74*, 4480-4490.

162. Dobson, C. M. (2002) Protein-misfolding diseases: Getting out of shape, *Nature 418*, 729-730.

163. Fink, A. L. (1998) Protein aggregation: Folding aggregates, inclusion bodies and amyloid, *Fold Des 3*, R9-R23.

164. Taylor, J. P., Hardy, J., and Fischbeck, K. H. (2002) Biomedicine - toxic proteins in neurodegenerative disease, *Science 296*, 1991-1995.

165. 2015 alzheimer's disease facts and figures, *Alzheimer's & Dementia: The Journal of the Alzheimer's Association 11*, 332-384.

166. Yon, J. M. (2001) Protein folding: A perspective for biology, medicine and biotechnology, *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.] 34*, 419-435.

167. Davis, C. M., Cooper, A. K., and Dyer, R. B. (2015) Fast helix formation in the b domain of protein a revealed by site-specific infrared probes, *Biochemistry-Us 54*, 1758-1766.

168. Nagarajan, S., Taskent-Sezgin, H., Parul, D., Carrico, I., Raleigh, D. P., and Dyer, R. B. (2011) Differential ordering of the protein backbone and side chains during protein folding revealed by site-specific recombinant infrared probes, *J Am Chem Soc 133*, 20335-20340.

169. Brewer, S. H., Song, B., Raleigh, D. P., and Dyer, R. B. (2007) Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy, *Biochemistry-Us 46*, 3279-3285.

170. Zimmermann, J., Thielges, M. C., Yu, W., Dawson, P. E., and Romesberg, F. E. (2011) Carbon-deuterium bonds as site-specific and nonperturbative probes for time-resolved studies of protein dynamics and folding, *J Phys Chem Lett 2*, 412-416.

171. Sato, S., and Fersht, A. R. (2007) Searching for multiple folding pathways of a nearly symmetrical protein: Temperature dependent phi-value analysis of the b domain of protein a, *J Mol Biol 372*, 254-267.

172. Sato, S., Religa, T. L., Daggett, V., and Fersht, A. R. (2004) Testing protein-folding simulations by experiment: B domain of protein a, *Proc Natl Acad Sci U S A 101*, 6952-6956.

173. Baxa, M. C., Freed, K. F., and Sosnick, T. R. (2008) Quantifying the structural requirements of the folding transition state of protein a and other systems, *J Mol Biol 381*, 1362-1381.

174. Dyer, R. B., and Vu, D. M. (2012) 3.3 fast events in protein folding, In *Comprehensive biophysics* (Egelman, E. H., Ed.), pp 34-42, Elsevier, Amsterdam.

175. Vu, D. M., Peterson, E. S., and Dyer, R. B. (2004) Experimental resolution of early steps in protein folding: Testing molecular dynamics simulations, *J Am Chem Soc 126*, 6546-6547.

176. Vu, D. M., Myers, J. K., Oas, T. G., and Dyer, R. B. (2004) Probing the folding and unfolding dynamics of secondary and tertiary structures in a three-helix bundle protein, *Biochemistry-Us 43*, 3582-3589.

177. Garcia, A. E., and Onuchic, J. N. (2003) Folding a protein in a computer: An atomic description of the folding/unfolding of protein a, *Proc Natl Acad Sci U S A 100*, 13898-13903.

178. Guo, Z., Brooks, C. L., 3rd, and Boczko, E. M. (1997) Exploring the folding free energy surface of a three-helix bundle protein, *Proc Natl Acad Sci U S A 94*, 10161-10166.

179. Cheng, S., Yang, Y., Wang, W., and Liu, H. (2005) Transition state ensemble for the folding of b domain of protein a: A comparison of distributed molecular dynamics simulations with experiments, *J Phys Chem B 109*, 23645-23654.

180. Alonso, D. O., and Daggett, V. (2000) Staphylococcal protein a: Unfolding pathways, unfolded states, and differences between the b and e domains, *Proc Natl Acad Sci U S A 97*, 133-138.

181. Kolinski, A., and Skolnick, J. (1994) Monte carlo simulations of protein folding. Ii. Application to protein a, rop, and crambin, *Proteins 18*, 353-366.

182. Ghosh, A., Elber, R., and Scheraga, H. A. (2002) An atomically detailed study of the folding pathways of protein a with the stochastic difference equation, *Proc Natl Acad Sci U S A 99*, 10394-10398.

183. Shao, Q., and Gao, Y. Q. (2011) The relative helix and hydrogen bond stability in the b domain of protein a as revealed by integrated tempering sampling molecular dynamics simulation, *The Journal of chemical physics 135*, 135102.

184. Shao, Q. (2014) Probing sequence dependence of folding pathway of alpha-helix bundle proteins through free energy landscape analysis, *J Phys Chem B 118*, 5891-5900.

185. Lei, H., Wu, C., Wang, Z. X., Zhou, Y., and Duan, Y. (2008) Folding processes of the b domain of protein a to the native state observed in all-atom ab initio folding simulations, *The Journal of chemical physics 128*, 235105.

186. Huang, F., Lerner, E., Sato, S., Amir, D., Haas, E., and Fersht, A. R. (2009) Time-resolved fluorescence resonance energy transfer study shows a compact denatured state of the b domain of protein a, *Biochemistry-Us 48*, 3468-3476.

187. Dimitriadis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G., and Smith, D. A. (2004) Microsecond folding dynamics of the f13w g29a mutant of the b domain of staphylococcal protein a by laser-induced temperature jump, *P Natl Acad Sci USA 101*, 3809-3814.

188. Arora, P., Oas, T. G., and Myers, J. K. (2004) Fast and faster: A designed variant of the b-domain of protein a folds in 3 mu sec, *Protein Sci 13*, 847-853.

189. Marecek, J., Song, B., Brewer, S., Belyea, J., Dyer, R. B., and Raleigh, D. P. (2007) A simple and economical method for the production of 13c,18o-labeled fmoc-

amino acids with high levels of enrichment: Applications to isotope-edited ir studies of proteins, *Organic letters 9*, 4935-4937.

190. Jackson, M., and Mantsch, H. H. (1991) Protein secondary structure from ft-ir spectroscopy - correlation with dihedral angles from 3-dimensional ramachandran plots, *Can J Chem 69*, 1639-1642.

191. Susi, H., Timasheff, S. N., and Stevens, L. (1967) Infrared spectra and protein conformations in aqueous solutions. I. The amide i band in h2o and d2o solutions, *J Biol Chem 242*, 5460-5466.

192. Huang, R., Kubelka, J., Barber-Armstrong, W., Silva, R. A., Decatur, S. M., and Keiderling, T. A. (2004) Nature of vibrational coupling in helical peptides: An isotopic labeling study, *J Am Chem Soc 126*, 2346-2354.

193. Turner, D. R., and Kubelka, J. (2007) Infrared and vibrational cd spectra of partially solvated alpha-helices: Dft-based simulations with explicit solvent, *J Phys Chem B 111*, 1834-1845.

194. Barber-Armstrong, W., Donaldson, T., Wijesooriya, H., Silva, R. A. G. D., and Decatur, S. M. (2004) Empirical relationships between isotope-edited ir spectra and helix geometry in model peptides, *J Am Chem Soc 126*, 2339-2345.

195. Myers, J. K., and Oas, T. G. (2001) Preorganized secondary structure as an important determinant of fast protein folding, *Nat Struct Biol 8*, 552-558.

196. Islam, S. A., Karplus, M., and Weaver, D. L. (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins, *J Mol Biol 318*, 199-215.

197. Jayachandran, G., Vishal, V., Garcia, A. E., and Pande, V. S. (2007) Local structure formation in simulations of two small proteins, *J Struct Biol 157*, 491-499.