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:

_____ _____

Amanda Katherine Cooper Date

Recombinant Site Specific Labeling of Proteins with IR Probes

By

Amanda Katherine Cooper

Masters of Science

Chemistry

_____

Dr. Richard Dyer

Advisor

_____

Dr. Vince Conticello

Committee Member

_____

Dr. Emily Weinert

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____ Date

Recombinant Site Specific Labeling of Proteins with IR Probes

By

Amanda Katherine Cooper
B.S., Georgia Institute of Technology, 2007

Advisor: Dr. R. Brian Dyer, Ph.D.

An abstract of
a thesis 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
Master of Science in Chemistry
2013

Abstract

Recombinant Site Specific Labeling of Proteins with IR Probes
By Amanda Katherine Cooper

The area of protein folding has a number of unanswered questions regarding the basic principles of how a protein achieves its native three dimensional folded conformation. Ultrafast folding proteins are small proteins with simple secondary structures that are models for kinetic studies that will provide more information on the fundamentals of how these structures form. One ultrafast folding model system with a high alpha helical content is the protein BdpA from staphylococcus. This protein has been studied both computationally and experimentally with conflicting results. The resolution of these contradictions through experiment will give insight into how alpha helical proteins form.

We have developed a labeling methodology for BdpA using a methionine auxotrophic cell line that allows for the incorporation of a heavy isotope infrared active label at a specific site of the protein. By constructing three mutant versions that incorporate this label in one of BdpA's three alpha helices, we expressed three proteins with a site in each helix for use in monitoring the kinetic parameters at the location of the label. Characterization of the proteins using infrared and circular dichroism spectroscopy shows that the point mutation introduced in different locations in each of the helices did not significantly destabilize the system, and that the label produced an isolated infrared signature that shows a temperature dependence that can be harnessed for kinetic measurements.

Preliminary kinetic measurements using temperature-jump infrared spectroscopy show that the label absorbance can be monitored for the derivation of kinetic parameters. Therefore, the insertion of these labels gives three different mutant versions of the protein with labels in various helical positions that, along with the wild type, can be compared in order to determine when the helices form in regards to the folding pathway of the protein. The derived kinetics can then be used to describe the folding pathway of the protein. By introducing minimally perturbing labels into various sites in BdpA, we can monitor site specifically using infrared spectroscopy the kinetics occurring at a highly localized site of the protein.

Recombinant Site Specific Labeling of Proteins with IR Probes

By

Amanda Katherine Cooper
B.S., Georgia Institute of Technology, 2007

Advisor: Dr. R. Brian Dyer, Ph.D.

A thesis 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
Master of Science in Chemistry
2013

Acknowledgements

I would first like to thank my advisor, Dr. Brian Dyer. Throughout these few years he has both encouraged my work in the laboratory and my interest in teaching. He was always available for help whenever my project was about the get the best of me. His direction and mentoring were invaluable. None of this would have been possible without the help of my advisor.

I would also like to thank my committee members Dr. Vincent Conticello and Dr. Emily Weinert for their help over the past few years. My project would not have been successful without their suggestions and guidance.

I would like to thank my lab mate Charu Kumar for teaching me everything I know about expressing proteins. I would also like to thank my lab mate and close friend Caitlin Davis for teaching me how to use almost all of the spectrometers that I used and helping me numerous times with analyzing my data. I also want to thank my other lab mates, both past and present for their help.

Most of all I would like to thank my wonderful parents Kay and Bill Cooper, my friends, the best of best friends Cole Welch, and my amazing boyfriend Kyle Digby. In the most stressful of times, these were the people who helped me through and kept my spirits up. Their support and love have been indescribable over the years and I cannot thank them enough for it.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BdpA	B domain of protein A
CD	Circular dichroism
Da	Daltons
DNA	Deoxyribonucleic acid
E. Coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
FTIR	Fourier Transform Infrared
ΔG	Gibbs free energy
H1	Helix 1
H2	Helix 2
H3	Helix 3
Hz	Hertz
His	Histidine
IR	Infrared
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kD	Kilodaltons
kHz	Kilohertz
LB solution	LB broth rich media solution
L	Liters
T_M	Melt temperature
µg	Microgram
µL	Microliter
µm	Micrometer
µs	Microsecond
mdeg	Millidegree
mg	Milligram
mL	Milliliter
mM	Millimolar
M	Molar
nm	Nanometers
ns	Nanoseconds
Nd-YAG	Neodymium-doped yttrium aluminum garnet
Ni	Nickel
OD	Optical density
QCL	Quantum cascade laser

s	Seconds
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
T-jump	Temperature-jump
t-RNA	t-Ribonucleic acid
v/v	Volume/volume
w/v	Weight/volume

Introduction

Protein Folding

Proteins are one of the most significant biomolecules present in a cell. In their capacity as enzymes, proteins catalyze the majority of reactions that occur in a cell. Proteins can also bind other molecules such as DNA, in which case they can regulate gene expression. In membranes, imbedded proteins create channels for the passage of nutrients into the cell or toxins out of the cell. These are only a few of the essential functions performed by protein molecules; however, none of these functions can be properly performed if a protein is not in its correct three-dimensional conformation. For proteins that adopt a highly structured conformation post expression this state is known as the folded state.[1] Understanding how proteins attain their highly specific folded state is a prominent area of focus in the study of cellular systems and has come to be known as the "protein folding problem".[2] Along the protein folding pathway of an unstructured protein to its native fold, there are a number of steps that can lead to off pathway

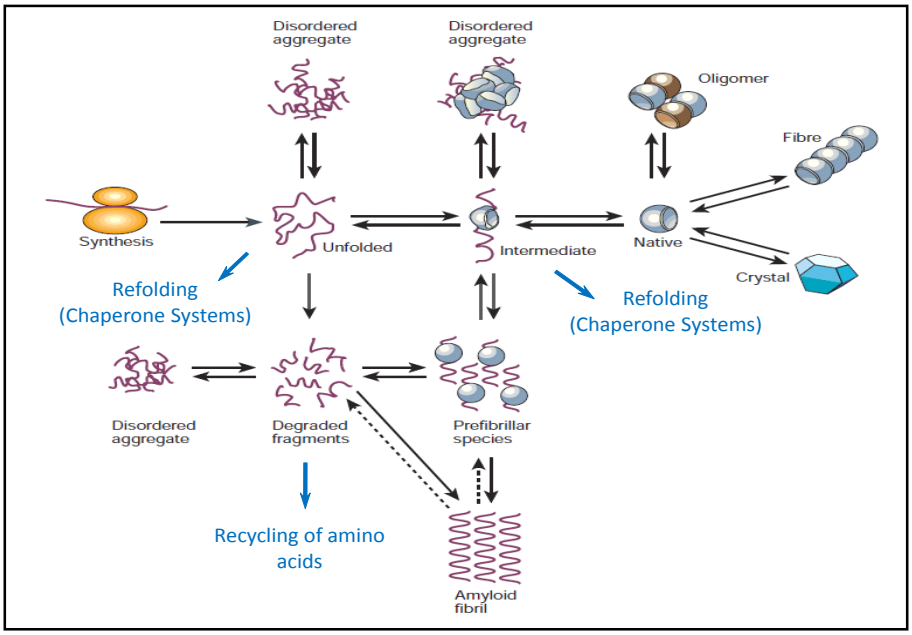


Figure 1: *Pictorial representation of protein folding reprinted with permission from C. M. Dobson, Nature* ***2003,*** *426, 884-890.*

misfolding problems including aggregation and fibril formation as seen in Figure 1. The cell has several rescue pathways to accommodate for the formation of these misfolded proteins including the use of chaperonin proteins for refolding and degradation of the misfolded products. However, these rescue pathways are not always able to correct the protein misfolding, leading to the accumulation of the misfolded products. Improper folding of proteins is now known to be the cause of a number of diseases. Misfolded proteins can form aggregates in diseases such as Parkinson's or completely lose their functionality resulting in a disease state as occurs in cystic fibrosis.[3-5]

In order to understand what leads to the misfolded state of a protein, it is important to first understand the basic principles underlying how a protein folds correctly. Even the fundamental properties of protein folding, including rate of secondary structure formation, amount of native like structure in the unfolded state and whether or not proteins follow a single folding pathway or have multiple possibilities for proceeding to the folded state, are still under debate.[1,6] Understanding the answers to these questions for simple protein systems will lead to the possible adaptation of these principles to proteins of greater complexity and for use in improving theorists' ability to correctly predict the folding pathway of a protein.

A protein's three dimensional structure is dependent on its amino acid sequence; however, much remains to be learned about how specific residues impact the final determination of a protein's structure and the pathway or pathways along which the protein achieves its final fold. The question of how proteins arrive in their final folded conformation has come to be known as the "protein folding problem." In 1968, Levinthal recognized that a protein cannot fold using a random sampling method due the large number of degrees of freedom for even a simple protein.[7,8] For a given protein, it can be estimated that the number of possible conformations is

around 3^N, where N is equal to the number of amino acids in the protein. The time required for a small protein of around 100 amino acids to sample every possible conformation would be longer than the age of the universe assuming one nanosecond as the timescale for the successful folding of a protein into one of the conformations.[9] Since proteins fold on a scale of microseconds for fast folders and seconds for slower folding proteins, there must be factors that narrow the portion of the possible phase space that a protein will sample along its folding pathway to its final biologically active conformation. This realization, also known as the Levinthal paradox, has led to an increasingly active interest in determining exactly how proteins are able to fold at such a rapid pace.

There are a number of proposed models for picturing the folding pathway of a protein.[3,10] The models focus on the resolution of the Levinthal paradox on the principle that the number of conformations available for a protein to sample is limited by energetic parameters. By mapping the energy landscape, possible folding pathways can be determined. The conformational energy is heavily biased towards the folded state which implies the protein only need sample a portion of the possible energy landscape putting folding back on the correct timescale. In monitoring protein folding, the energy landscape needs to be determined, including any important features such as energetic traps or minima from possible on or off pathway intermediates, bottlenecks where the number of available conformations for sampling drastically reduces in a short period of time, and other features in the energy landscape. One popular depiction of the protein folding energy landscape is the protein folding funnel.[1,3] A pictorial representation of the folding funnel is shown in Figure 2. In the funnel model, the number of possible conformations at any given point in a protein's folding pathway is represented by the diameter of the funnel corresponding to that point in the pathway. The funnel shape will depend on the energy landscape of the folding process. Folding proteins can go through various on or off path

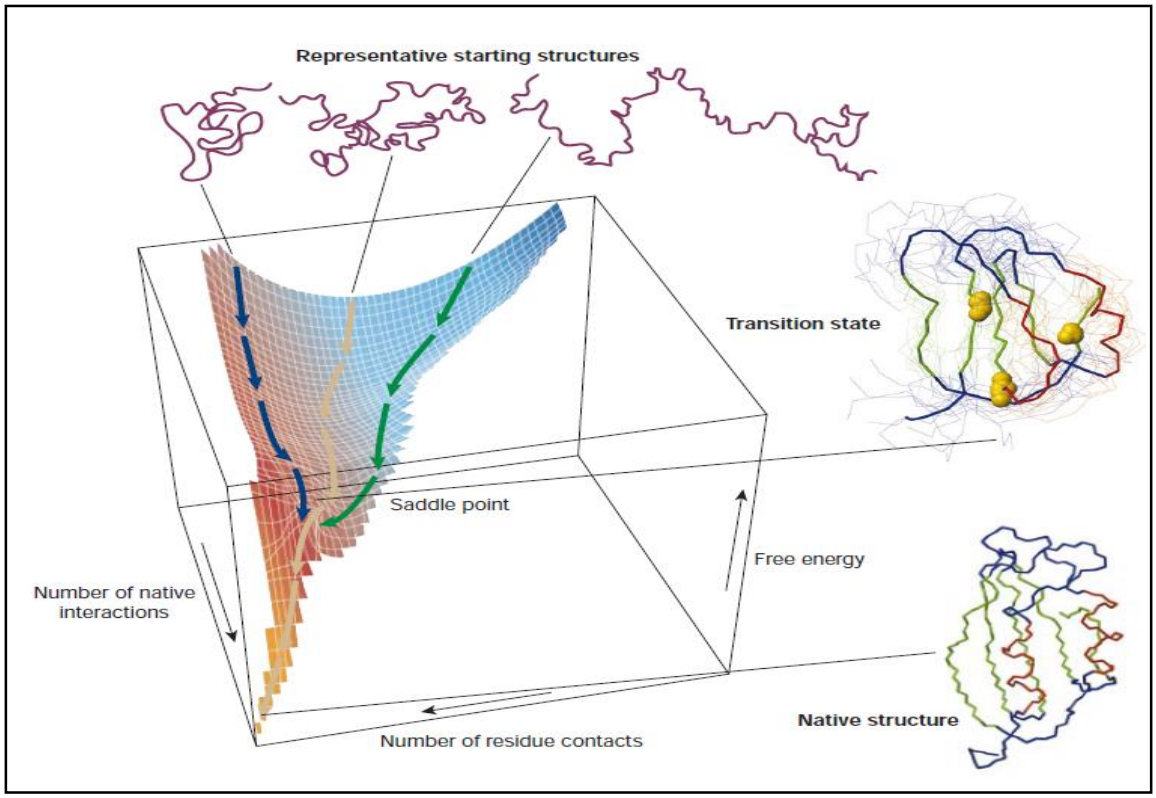


Figure 2: *Pictorial representation of protein folding as a 'folding funnel' reprinted with permission C. M. Dobson, Nature 2003, 426, 884-890.*

intermediates and possibly follow multiple pathways to reach their final folded state shown as energy minima in funnel models. However, the fact remains that the number of pathways possible must be limited for a protein to reach its appropriate folded conformation on a realistic timescale.

The pathway through which a protein folds can be determined both experimentally and through simulation. Experiments can monitor the pathway through direct observation of labels or other means of observing the protein kinetics. However, folding pathways can be altered by the observation method used. Simulation can be used to monitor the folding of the protein without having to introduce any perturbations into the protein's folding pathway, yet current simulation methods still need further refinement in order to reproducibly and accurately predict a folding pathway. Therefore, it is essential to employ both experimental and theoretical techniques in

order to improve the available technology and better understand the fundamentals of protein folding.

In finding proteins to use as model systems for computer simulations, one of the most important characteristics necessary is simplicity. While computer simulations have seen immense improvements in the past few decades, there is still an exponential increase in time required for a simulation when compared to the length of a protein.[11-14] The majority of proteins fold on a timescale of microsecond to seconds. However, all atom simulations, which are desired for the great detail that can be achieved with them, are limited to the timescale of nanoseconds to microseconds.[15,16] Proteins that fold on these timescales are termed ultrafast folding proteins. Ultrafast folding proteins generally follow single exponential kinetics, although some have been shown to fold through obligatory intermediates.[17] The simplicity of the proteins and the kinetic pathway they follow allow for the use of all atom simulations to predict their folding pathways. Since experimental procedures have now been developed that allow for following relaxation kinetics on the microsecond and nanosecond timescale, these proteins are ideal to test the ability of simulation programs to emulate experimental data.

The simplistic structures of ultrafast folding proteins give information on the fundamental principles that underlie protein folding. Characteristic properties of a specific secondary structure such as alpha helices can be derived by analyzing protein with simple secondary structures. Comparing results from experiments involving different proteins with high similarity in their secondary structure gives information on how tertiary interactions and side chain organization can factor into the folding pathway of a protein. Finally, the refinement of these techniques will lead to the ability to add complexity into the system and analyze the folding of proteins in membranes and other more intricate systems. Site specific analyses introduce the

capability to examine the kinetics of a localized region, giving detailed information of what occurs along the pathway of the folding of a protein.

BdpA Background

The ultrafast folding B domain of protein A (BdpA) has been previously established as a candidate for kinetic studies.[6,10,18-32] BdpA is an immunoglobulin-binding domain found on the surface of *staphylococcus aureus*.[33] By binding an immunoglobulin on a host cell, engulfment of the cell by phagocytes is prevented. The domain is a three helix protein of 60 amino acids. The structure is shown in Figure 3. The three helices are numbered 1-3 starting at the N terminus. Helix one (H1) contains residues 10-19, turn one residues 20-24, helix two (H2) residues 25-37, turn two residues 38-41 and helix three (H3) residues 42-58. The sequence can be seen in Figure 4 including the location of the helices indicated by the orange loops. The relaxation kinetics of BdpA occurs on a micro to nanosecond time scale, qualifying it as an ultrafast folding protein.[34] The wild type also shows a significant temperature stability with a T_M of 73.3 ± 1°C and is not prone to aggregation even at higher temperatures.

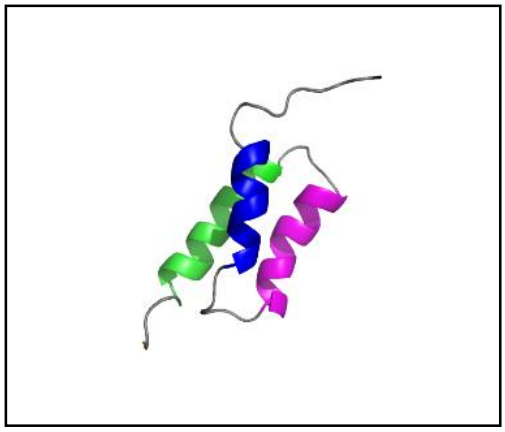


Figure 3: *BdpA structure from www.pdb.org PDB ID 1BDD*

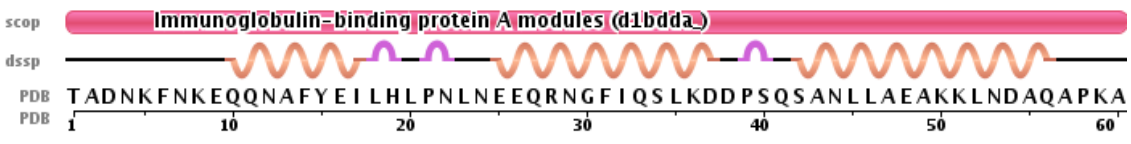


Figure 4: *BdpA sequence from www.pdb.org PDB ID 1BDD*

The fast folding rate and simple structure of BdpA make it a great candidate for computer simulations. A number of simulations have been performed on the protein and have given varying results as to the predicted pathway that BdpA follows in achieving its final folded conformation. These results have also been compared to experimental studies with varying success.[6,10,18-26] Experimental data have indicated that the folding transition state includes a nearly fully formed helix two and a partially formed helix one. Some computational studies support these findings, although they usually predict a larger amount of helix formation in helix one than as shown by experimental findings. Other computational studies show contradicting results with initial formation of helix three. Additional experimentation can be used to support one of the proposed pathways and thereby help with the refinement of the simulation of the folding pathway of BdpA as well as giving information on the basic principles of the folding of alpha helical proteins.

Reported Computational Results

BdpA is one of the most widely simulated proteins in the field of protein folding; however, the simulations do not agree with the folding pathway as determined by experiment or even with one another. It has been proposed that the BdpA transition state includes a nearly fully formed H2, partially formed H1 and interactions between H1 and H2. There is also additional structure at the turn between H2 and H3 and at the N-terminus. This proposal is based on previous experimental evidence including an extensive phi-value analysis which analyzes the changes in free energy caused by altering specific residues in the transition state of the protein. Computational methods, however, predict the formation of H1 in the transition state.

Understanding the relaxation kinetics of BdpA will allow for comparisons with computational models. The experimental data can be used to improve the computational methods employed. Kolinski and Skolnick performed one of the first computational studies using a Monte Carlo approach.[27] Their results found that folding followed a pathway that initiates with the formation of the C-terminal hairpin. Ghosh *et al* used a stochastic difference equation method and also found initial formation of the C-terminal end, namely H3.[28] Daggett *et al* used all-atom temperature melt molecular dynamics and concluded that the first helix to be unfolded was H1, followed by H2 and lastly H3.[24] These studies are just a few examples of the collective simulations that indicate greater stability and initial formation of H3 in apparent contradiction to the folding pathway predicted by experiment.

Another group of simulation studies have found results that emulate those found by experiment. Brooks et al used umbrella-sampling and explicit solvent molecular dynamics and found that H1 and H2 formed first in the simulated folding pathway followed by formation of H3.[29] Onuchic *et al* used replica exchange molecular dynamics to model the folding pathway and also found that H1 and H2 formed a core in the transition state that H3 later docks to.[25] Liu *et al* used a distributed computing method that looked at systems with topological similarity to the native fold.[30] The results were similar to the first two with substantial formation of H2 with docking of H1 followed later by docking of H3. These computational studies and others seem to support the pathway derived from the majority of experimental results. The disagreement in the proposed pathways indicates that experimental evidence is required to resolve the differences that are inherent in the various methodologies used in performing computational folding studies.

Reported Experimental Results

Experimental studies that have been conducted on BdpA propose a folding pathway that includes a transition or intermediate state involving a nearly fully formed H2 with tertiary contacts to a partially formed H1 as the primary structural elements present. An early experiment conducted by Bai *et al* aimed at finding any possible intermediates along the folding pathway was unable to conclusively find experimental evidence of an intermediate.[20] However, their experiment focused on a mutant with a destabilized H1 due to indication by hydrogen-exchange that any possible intermediate would contain a structure focused on H2 and H3. The experiments were also unable to obtain the necessary time resolution requirements to observe kinetic phases faster than microseconds. A later experiment conducted by Fersht et al used an extensive phi-value analysis.[18,19,31] Phi-value analysis employs mutations located at various sites in a protein's sequence chosen to interrupt specific interactions or destabilize helix formation without perturbing other segments of the protein structure. By determining the change in Gibb's free energy of the transition and comparing it to the change in Gibb's free energy of the final product, a phi-value can be determined. This will show the extent to which the mutated residue contributes to the native structure of the protein. High phi-values indicate a native like interaction for the residue, implying that the native interaction has formed in the transition state. In the phi-value analysis of BdpA, high phi-values were obtained for H2 primarily when the secondary structure formation was probed. Additional helicity was observed in H1, but for fewer residues. In the analysis of tertiary interactions, high phi-values were obtained for residues that correlate to interactions between H1 and H2. Sosnick *et al* used psi-anaylsis to evaluate the folding pathway of BdpA.[21] Psi-analysis is similar to phi-value analysis, but uses bihistidine metal ion binding sites to stabilize transition states and obtain values on the distance

between the two partners. In the psi-analysis, the most present helix in the transition state was H2. A recent study conducted by Dyer *et al* used fluorescent probes coupled with T-jump fluorescence spectroscopy to monitor kinetic states in the protein.[32] Three mutants were created, each including a fluorescent probe on one of the helices. Two kinetic phases were observed in the mutants, a fast phase corresponding to helix formation and a slower phase corresponding to the global folded state formation. The only mutant to show the fast kinetics phase was the mutant containing a fluorescent probe on H2. These results further support the proposed folding pathway. However, insertion of the fluorescent probes can potentially affect the folding pathway of the protein. Therefore, additional experimentation using less perturbing labels can be used to test this hypothetical pathway.

Infrared Spectroscopy for Monitoring Folding

Infrared spectroscopy is rapidly growing in popularity as a useful tool in the analysis of protein structures and folding pathways.[35-42] The technique is highly sensitive to secondary structure and can give specific information about the presence of secondary features in the protein structure. Another form of spectroscopy that can give secondary structure information is circular dichroism (CD) spectroscopy; however, CD spectroscopy is limited in its ability to resolve similar secondary features as well as its ability to differentiate between secondary structure changes such as solvation or isotope labeling.[43-47] Infrared spectroscopy is able to easily resolve differences in absorbances between features such as beta turn and beta sheets and can differentiate between the solvation state of alpha helices. A table showing the observed frequencies for various secondary structures can be seen in Table 1. The band described in this

Confirmation	Mode	D_2O	H_2O	Solid
Antiparallel-chain pleated sheet	$\nu(\pi,0)_A$	1632	1632	1632
	$\nu(0,\pi)_A$	1675	1690	1690
α Helix	$\nu(0)_\alpha$	1649		
		1650	1652	1652
Unordered	$\nu(u)_A$	1643	1656	
		1643	1656	

Table 1: *Table showing absorbances of various secondary structures.*[43]

table is conventionally called the amide I band. The amide I band corresponds to the stretching frequency of the carbonyl groups in the backbone of the protein. When the protein is exchanged into D_2O, the band is called the amide I' band. The peaks described in the table are typically well resolved enough to be able to monitor several secondary structures simultaneously if there are multiple features present in the protein.

Another benefit of using infrared spectroscopy is the ability to determine if the secondary structure is exposed to solvent or buried within the protein. The hydrogen bonding of residues exposed to the solvent differs from the hydrogen bonding of residues within the protein interior causing a shift in the vibrational frequency.[34] The shift is large enough to be able to differentiate between a solvent exposed and a solvent buried feature. The resolved peaks can be followed over a temperature dependent denaturation in order to determine the thermodynamic stability of the protein. The frequencies can also be monitored using relaxation techniques to derive the kinetics of each of the different features. A general folding pathway can be determined from the kinetics derived from these relaxation measurements.

In order to determine a more detailed folding pathway, infrared active labels can be inserted into the protein at a given position.[32,48-50] Infrared active labels are typically much smaller than labels introduced for use in other spectroscopic methods. For the amide I' band, an ideal label is

a heavy isotope label of the carbonyl group. The labeled band will be shifted to a significantly lower frequency allowing for resolution of the labeled band from the unlabeled stretch absorbances.[51] The labeled frequency can then be monitored using relaxation techniques in order to determine the kinetics of folding at that specific location in the protein. This process can be used to determine kinetic rates site specifically in a protein. The ability to monitor a specific position in the protein sequence makes infrared spectroscopy an incredibly powerful technique in the determination of a protein's folding pathway.

Specific Aims

Specific Aim 1: Develop Expression Systems for Wild Type and Mutant Proteins

To examine the secondary structural kinetics of an alpha helical protein, the sixty amino acid B domain of protein A (BdpA) was analyzed using infrared spectroscopy to monitor site specific IR labels. An expression system for the wild type protein was first established using *E. coli* cultures in LB broth media. The infrared label mutants were then constructed using a combination of recombinant genetics and auxotrophic cell lines to insert a $^{13}C=^{18}O$ backbone isotope labeled methionine into each of the helices. The expression products were characterized by infrared spectroscopy to refine the protocol for successfully isolating all four proteins for use in infrared measurements.

Specific Aim 2: Determine Stability and Thermodynamic Parameters of Proteins

The isolated proteins were analyzed using circular dichroism and infrared spectroscopies to determine the thermodynamic stabilities of each of the proteins. Circular dichroism was used to determine if there were any changes in the secondary structure of the protein from the

mutations. The loss of helicity over a range of temperatures was then monitored using temperature dependent circular dichroism. From the change in helicity, a melt temperature was determined for the protein corresponding to a 50% concentration of folded vs. unfolded protein. Infrared characterization was also performed to ensure that the protein label was resolved from absorptions of the unlabeled backbone vibrations. Temperature dependent infrared spectroscopy was used to monitor changes in the protein absorption and determine melt temperatures using a second spectroscopy. By comparing the melt temperatures determined from the various methods, the thermodynamic stability of the various mutant proteins was related to the stability of the wild type protein. These melt temperature values were also used to determine the Gibbs free energy of each protein for comparison of any alterations to the thermodynamics of BdpA caused by the introduction of the mutations.

Specific Aim 3: Probe kinetics of structure formation in an alpha helical ultrafast folding protein

Preliminary kinetic studies were performed on the wild type protein and helix three mutant by following the change in infrared absorbances for the wavelength of the labeled backbone vibration and unlabeled backbone vibrations using temperature-jump infrared spectroscopy. The preliminary data showed that both the labeled and unlabeled absorbances were observable using this technique. Therefore, the kinetic data for each helix can be obtained by using temperature-jump spectroscopy to monitor the kinetics of each of the mutant proteins. By monitoring the relaxation kinetics of the protein backbone, relaxation rates can be obtained for the formation of secondary structure for each of the individual helices and an overall picture of the formation of secondary structure can be determined. The obtained rates can then be

compared to those obtained in previous studies, both computational and experimental, and to the rates obtained by labeling the backbone.

Chapter 1 Protein Expression, Purification and Characterization

Introduction

In order for a labeling technique to be effective for use in any sort of analysis, it is essential to develop a reliable methodology for preparing the labeled protein with a significant yield and high labeling efficiency. Yields from synthetic protein methods lose efficiency for each amino acid in a protein sequence, making it unfeasible for synthesizing larger protein.[52] For example, BdpA with 60 amino acids would have a theoretical yield of 54% if the yield of each coupling step was 99%. This coupling efficiency is not achievable for each amino acid coupling, meaning that the actual synthetic yield of BdpA would be even lower. Therefore, we focused on the use of an expression system for producing the wild type and various labeled mutant versions of BdpA used in this study. Our label of interest incorporates a heavy isotope $^{13}C=^{18}O$ carbonyl bond into the backbone of the protein at a specific site in the sequence. The location where the label was inserted was selected by introducing a methionine mutation at the desired position. The natural t-RNA synthetase pathway of *E. coli* was then harnessed to introduce the heavy isotope methionine at the site of the mutation.[53]

In order to introduce the labeled methionine to the sequence during the expression phase, an auxotrophic cell line was used. The cell line is unable to synthesize methionine and requires a supplement in order to express any methionine containing proteins. By supplementing a labeled version of methionine at the time of induction, the methionine that is inserted into the overexpressed protein will contain the desired label. In the case of a heavy isotope label, the exchanged amino acid was added to the minimal media used and the label was incorporated into the protein without reexchange of the labeled oxygens. This method also has the potential

for use in the introduction of any methionine based label that is recognized by the t-RNA synthetase.

Any mutation introduced into a protein sequence has the potential to destabilize the protein structure. Methionine was selected to avoid introducing additional mutations into the protein sequence since methionine is the rarest amino acid. In BdpA there are naturally no methionines so no additional mutations were necessary. By minimizing the number of mutations to the protein, we also minimized the potential for destabilizing the protein. In terms of stability, the infrared sensitive heavy isotope label presents a significant advantage over other labeling techniques by being non-perturbing to the mutant structure. The only concern for protein stability lies in the introduction of the methionine mutation. The effect of this mutation is seen in the ability of the expression system to overexpress the mutant in comparison to the wild type, using spectroscopy to monitor differences in secondary structure, and determining the thermodynamic stability. The mutant expression systems gave comparable yields to the wild type protein and the secondary structure of the mutants showed little change when examined by circular dichroism and infrared spectroscopy. Thermodynamic studies also showed comparable stabilities, and will be shown in the following chapter. These results combined show that the insertion of the labeled methionine does not significantly alter the structure of the protein.

The development of this expression system for BdpA allows for the site-specific introduction of an infrared active label at a chosen position in the sequence. Since we are interested in the kinetics of each of the three helices, we designed three mutant versions of BdpA, each with a single label in one of the helices. Wild type protein expression was developed in an LB broth *E. coli* system, while the mutant protein systems utilized minimal media to introduce the labeled

methionine to the protein sequence. We were able to design expression systems and successfully isolate all four of the desired proteins.

Materials and Methods

Construction of Plasmids

The gene coding sequence for BdpA was incorporated into a pET-30a(+) vector along with a coding sequence for the addition of a histidine affinity tag and TEV protease cleavage site to create the plasmid for the wild type. The vector incorporates kanamycin resistance for use in selection. The mutant plasmids were constructed in house by introducing a point mutation at the selected site. The mutant plasmids were sequenced (Beckman Coulter, Indianapolis, IN) to ensure that the sequence was correctly mutated. The plasmids were introduced into a chemically competent DH5Lα cell line using a heat shock method. The cells were then grown in LB broth rich media solution (LB solution) and selected using LB/agar plates with 100 µg/mL kanamycin. Selected cells were then grown in LB solution with 100 µg/mL kanamycin for plasmid amplification. The plasmids were isolated using the Qiagen QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) following the tabletop centrifuge protocol. Successful isolation of the plasmid was confirmed by agarose gel as shown in Figure 5. Isolated plasmids were stored at -20°C.

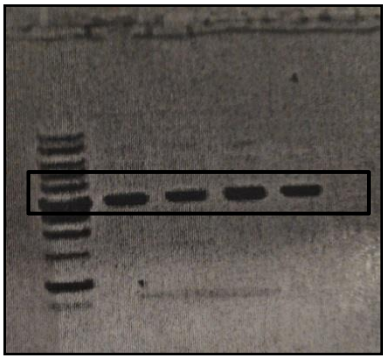


Figure 5: *Agarose gel of BdpA plasmid. Lane one shows markers, subsequent lanes show isolated plasmids. Boxed area shows plasmid near 5000 kD marker corresponding to approximate mass of BdpA plasmids..*

Cell Selection

The plasmids were introduced into a B843 pLysS cell line (Novagen, Foster City, CA) using heat shock. Cells containing the plasmid were selected using LB solution with 100 µg/mL kanamycin and 40 µg/mL chloramphenicol. Further selection was performed by plating the cultures on LB/agar plates with the same concentration of antibiotics. The plates were stored at 4°C. Glycerol cell stocks were also constructed from the cell culture media using a 50% mixture of sterilized 70% glycerol and cell culture. The glycerol stocks were stored at -80°C.

Wild Type Expression and Purification

Wild type cell cultures were grown using cells from the LB/agar plate in 100 mL LB solution at 37°C overnight. The cultures were then grown to an OD of 0.6-0.8 in LB solution at 37°C and induced with 1 mM IPTG. Protein was expressed for 24 hours at the same temperature. Samples were taken pre- and post-expression to determine the presence of expressed protein by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS PAGE) as seen in Figure 6. The proteins can be observed in the gel as a stronger band in the post expression sample below the 10kD maker, indicative of the ~8600 Da size of all four BdpA proteins. Fast protein liquid chromatography (FPLC) was used to isolate the expressed. Cells were harvested by centrifugation and resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) with a ratio of 10 mL of buffer to 1 L of culture harvested. Lysozyme

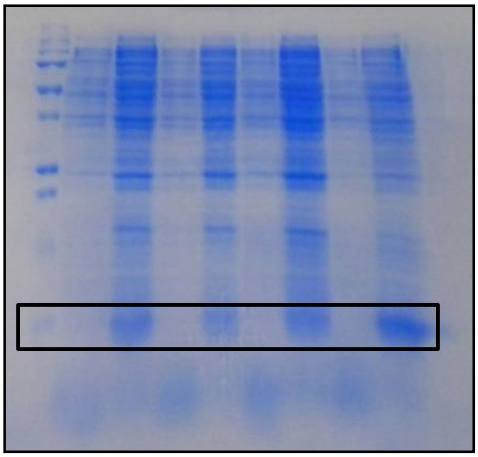


Figure 6: *SDS PAGE gel showing expression of BdpA. First lane shows markers, subsequent lanes are pre- and post-expression. Strong post expression band can be observed below 10 kD marker as shown in boxed area.*

was then added in a concentration of 1 mg per 1 mL of lysis buffer by dissolving in water and filtering. The cells were lysed by 30 minutes of stirring with the lysozyme followed by 30 minutes of sonication. The cell debris was pelleted out using centrifugation and the supernatant was collected and filtered. The filtered supernatant was then subjected to FPLC. A typical FPLC spectrum can be seen in Figure 7. The sample was loaded onto a Ni His prep FF 16/10 column (GE Healthcare Life Sciences, Pittsburgh, PA). Impurities were washed from the column using a wash buffer (20 mM NaH_2PO_4, 500 mM NaCl, 30 mM imidazole, pH 7.4). The sample was then eluted using a gradient protocol. The gradient ran from 0% to 100% elution buffer (20 mM NaH_2PO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4) with the second solvent being the wash buffer. BdpA samples generally eluted at a concentration of around 75% elution buffer. The collected fraction was then dialyzed into water. The dialyzed sample was lyophilized overnight to obtain pure, water free protein sample. All lyophilized protein samples were stored in septa capped eppendorf tubes in a desiccator.

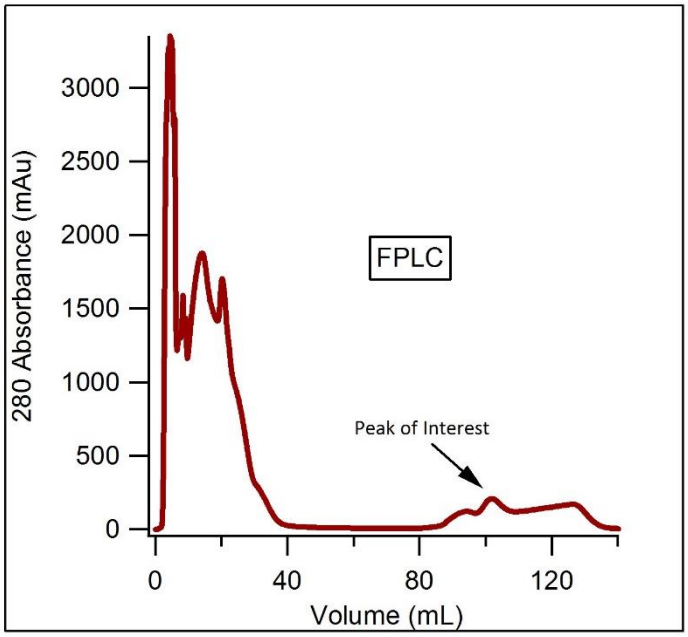


Figure 7: *Sample FPLC of BdpA. Second peak post sample wash, as indicated by the arrow was collected peak containing protein. Increasing imidazole concentration causes increasing baseline observed during elution of sample.*

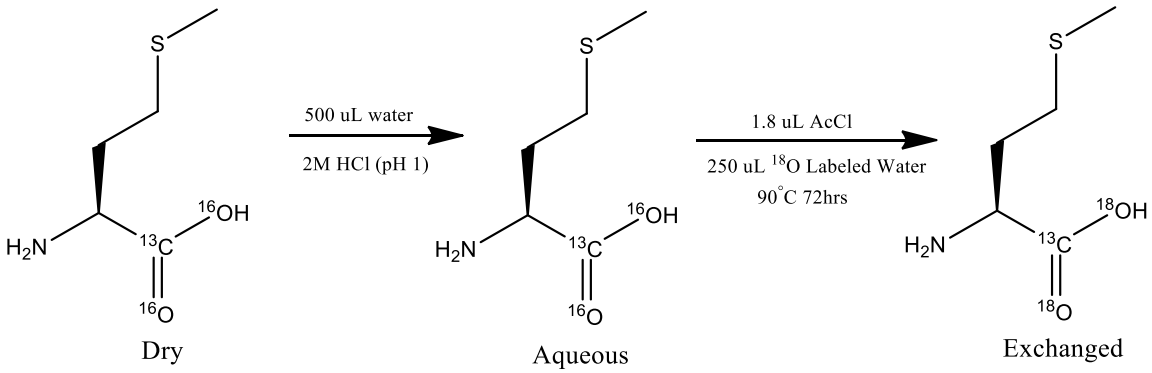


Figure 8: *Scheme showing exchange of ^{13}C labeled methionine to ^{18}O and ^{13}C labeled methionine*

Methionine Enrichment

Prior to expression in minimal media, ^{13}C labeled methionine was enriched to a ^{13}C=^{18}O labeled methionine using ^{18}O labeled water.[48] A 50 mg sample of ^{13}C labeled methionine (Cambridge Isotopes Tewksbury, MA) was dissolved in water and the pH of the solution was slowly adjusted to 1 using 2 M HCl. The highly acidic sample was then lyophilized overnight. A solution of 250

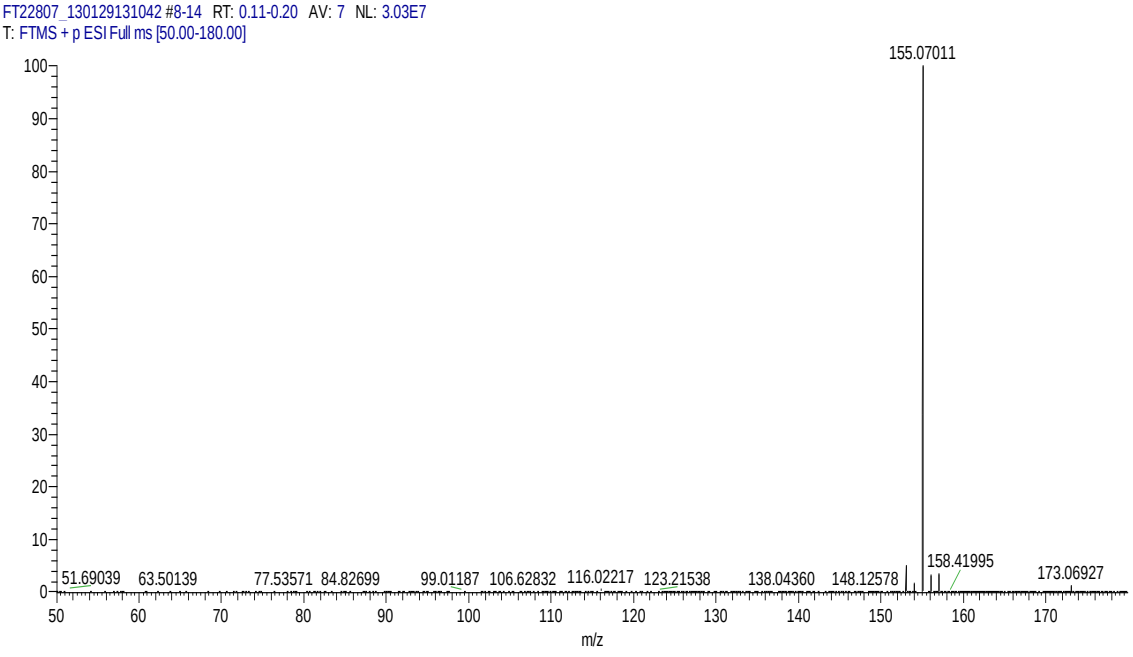


Figure 9: *Mass spectrum of labeled methionine sample. 155.07011 mass peak is protonated double oxygen labeled sample. Single oxygen labeled sample also appears in less than a 10% ratio. All samples were exchanged to this purity.*

µL of ^{18}O labeled water (Cambridge Isotopes Tewksbury, MA) and 1.8 µL dry acetyl chloride was prepared and the dried methionine was added to the solution. The methionine was allowed to exchange at 90°C for 72 hours in order to obtain complete exchange of both oxygens in the carboxylic acid functional group. An overview of this process can be seen in Figure 8. After completion of the exchange, a sample was submitted for high accuracy mass spectrometry to confirm the completion of the exchange. A sample mass spectrum for a labeled sample of methionine is shown in Figure 9 and characterization using infrared spectroscopy is shown in

Peak	^{13}C Labeled (cm^{-1})	^{13}C ^{18}O Labeled (cm^{-1})	Peak Shift (cm^{-1})
1	1310.04	1293.89	16.15
2	1345.89	1343.9	1.99
3	1393.27	1386.11	7.16
4	1445.77	1444.05	1.72
5	1504.03	1490.16	13.87
6	1545.64	1535.09	10.55
7	1616.06	1628.49	-12.43

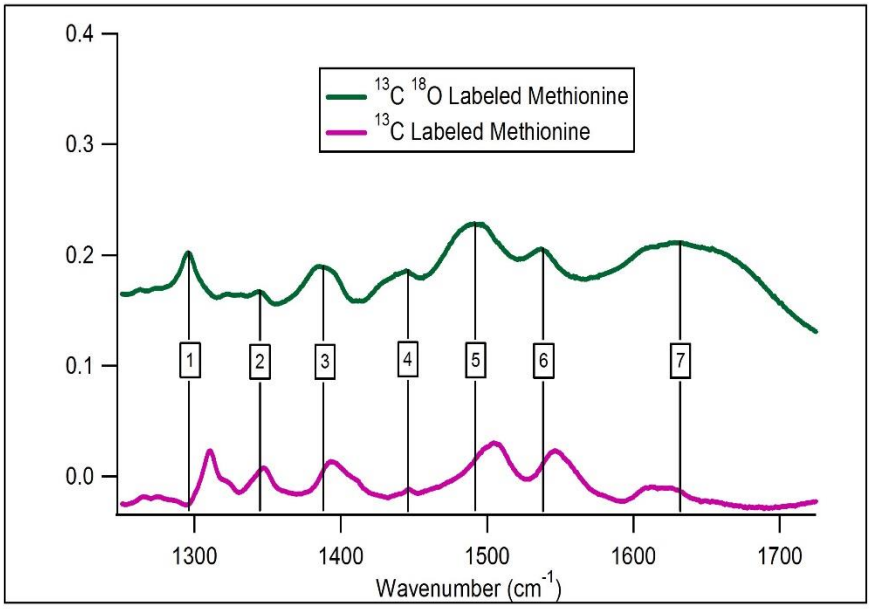


Table 2: *Tabulated data of IR peaks for ^{13}C only and ^{13}C ^{18}O labeled methionine and induced peak shift from addition of heavy oxygen*

Figure 10: *IR spectrum showing shifts of methionine peaks with introduction of ^{18}O*

Figure 10. Exchanged samples were then lyophilized again to remove any residual water and stored in an eppendorf tube with a septum in a desiccator to prevent the introduction of ^{16}O water and possible reexchange.

Mutant Expression and Purification

Minimal media expression systems were used to introduce the labeled methionine into the protein for each of the three mutants. An overnight culture was grown using 100 mL of LB solution with 100 µg/mL kanamycin and 40 µg/mL chloramphenicol for each desired liter of expression culture. The following day, 1 L of minimal media was made for each of the 100 mL cultures from the previous day. The 1 L of minimal media was comprised of 200 mL of 5x solution (0.11 M Na_2HPO_4, 0.022 M KH_2PO_4, 0.0086 M NaCl and 0.019 M NH_4Cl), 2 mL of 1 M $MgCl_2$, 100 µL of 1 M $CaCl_2$, 20 mL 20% glucose w/v, and 40 mg of each amino acid including unlabeled methionine. The remaining volume was comprised of water and all components of the minimal media were purified through either autoclaving or filtration. The cell culture was harvested using centrifugation and resuspended in the minimal media solution. The culture was allowed to grow to an OD of 0.6-0.8 and then harvested again using centrifugation. The cells were then resuspended in a newly constructed minimal media solution containing the same components, but with labeled methionine in place of the unlabeled methionine. Once the cell pellets were gently resuspended, the sample was induced using 1 mM IPTG and expression was allowed to proceed for 24 hours at 37°C. After the expression was complete, the protein was harvested using the same protocol as that outlined for the wild type protein with a typical yield of 5-10 mg per liter of expression. Successful expression was confirmed using SDS PAGE and mass spectrometry.

Removal of Hexa-Histidine Affinity Tag

The isolated proteins containing the histidine purification tag were subjected to a cleavage protocol using AcTEV protease (Invitrogen, Grand Island, NY). The protein was cleaved in batches of a maximum of 5 mg with a ratio of 12.4 units protease to 1 mg of protein. The cleavage buffer for a 5 mg sample of protein was comprised of 18.75 µL 0.1 M DTT, 93.75 µL AcTEV protease buffer (1 M Tris-HCl pH 8.0 and 10 mM EDTA), 6.2 µL 10 units/µL AcTEV protease in solution (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 5mM DTT, 50% (v/v) glycerol, 0.1% (w/v) Triton X-100) and 367.5 µL of water for a total volume of 500 µL of cleavage reaction. The reaction was conducted for 48 hours at 4°C. After the tag had been removed, the cleaved protein was separated from free tag using a Ni sepharose resin in a batch protocol. Aliquots of 500 µL of Ni Sepharose High Performance resin (GE Healthcare Life Sciences, Uppsala, Sweden) were first primed for batching by centrifugation and removal of the supernatant. The resin was then washed with an imidazole buffer (20 mM NaH_2PO_4, 500 mM NaCl, 30 mM imidazole, pH 7.4) to prevent background binding of non-tagged proteins. The reaction solution was introduced in a ratio of 100 µL solution to 500 µL primed resin. The protease includes a histidine affinity tag and therefore bound the resin along with the removed histidine tag from the protein. The mixture was mixed overnight on a rotisserie and the supernatant harvested the next day. 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 then dialyzed into water overnight to remove the NiEDTA product and the additional components of the cleavage buffer solution. The dialyzed sample was lyophilized overnight to obtain pure, water free protein sample. All lyophilized protein samples were stored in septa capped eppendorf tubes in a desiccator.

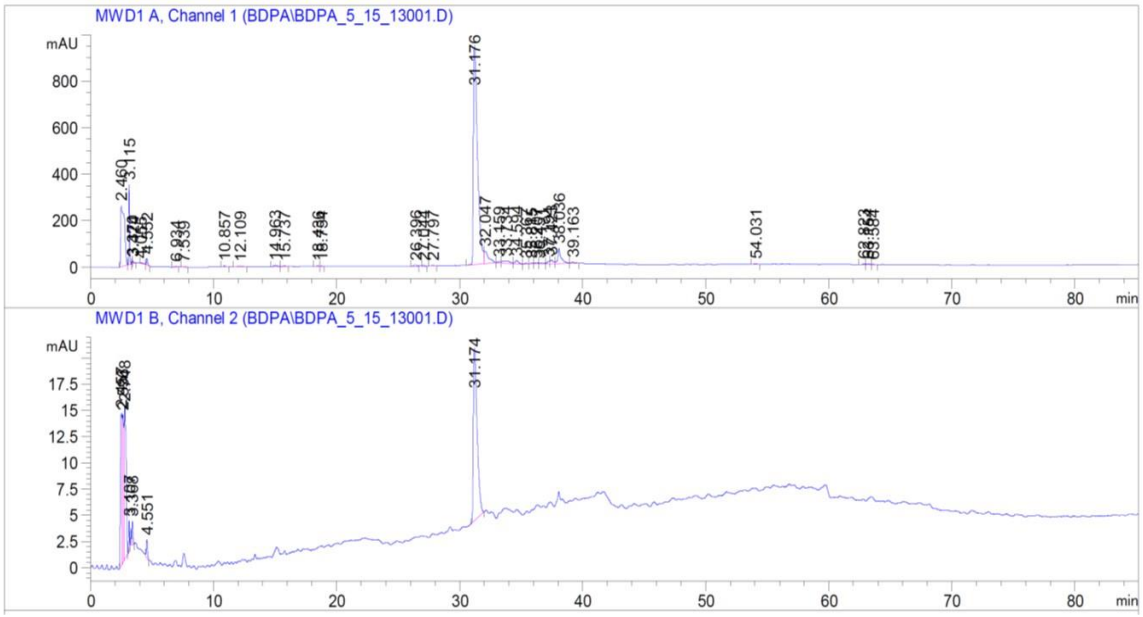


Figure 11: *HPLC spectrum of BdpA sample. Samples appeared around 30-35% ACN with few additional peaks seen along the gradient.*

Further purification was necessary for some proteins and achieved using high performance liquid chromatography (HPLC) purification on an Agilent 1000 system (Agilent, Santa Clara, CA). The samples were dissolved in water to 2 mL and injected onto an Agilent Zorbax 300SB-C3 column. The absorbance was monitored at 280 nm and 222 nm. Prior to injection the column was washed with a 90:10 mixture of water with 5 mM HCl to acetonitrile with 5 mM HCl. Once a baseline was established, the sample was injected and a gradient protocol was followed. The gradient ran from 10% acetonitrile with 5 mM HCl to 100% acetonitrile with 5 mM HCl with water with 5 mM HCl being the second solvent. BdpA peaks consistently appeared around 30-35% acetonitrile as shown in the sample HPLC in Figure 11. The peak was collected and subjected to dialysis overnight. The dialyzed sample was then lyophilized overnight to obtain pure, water free protein sample. All lyophilized protein samples were stored in septa capped eppendorf tubes in a desiccator.

Analysis by IR spectroscopy

Once each protein was synthesized and characterized through biochemical techniques, further characterization was performed using FTIR spectroscopy. BdpA wild type protein shows two structured peaks at 1632 cm^{-1} and 1648 cm^{-1} when it is folded and a broad unstructured feature at higher temperatures around 1670 cm^{-1}.[34] The heavy isotope label mutants also have an additional peak at around 1560 cm^{-1} corresponding to the label. It is therefore important to confirm that there are no contaminating peaks in any of these regions before using the samples to perform spectroscopy. The characterization of BdpA by temperature dependent IR can be seen in Figure 12. The spectra that are plotted show features that are changing with temperature. These features are difficult to observe in the total absorbance spectra, but can clearly be observed by subtracting the lowest temperature spectrum from each subsequent temperature spectrum to construct difference spectra. The two dashed lines show the location of the two structured protein peaks at 1632 cm^{-1} and 1648 cm^{-1} for comparison between the various spectra.

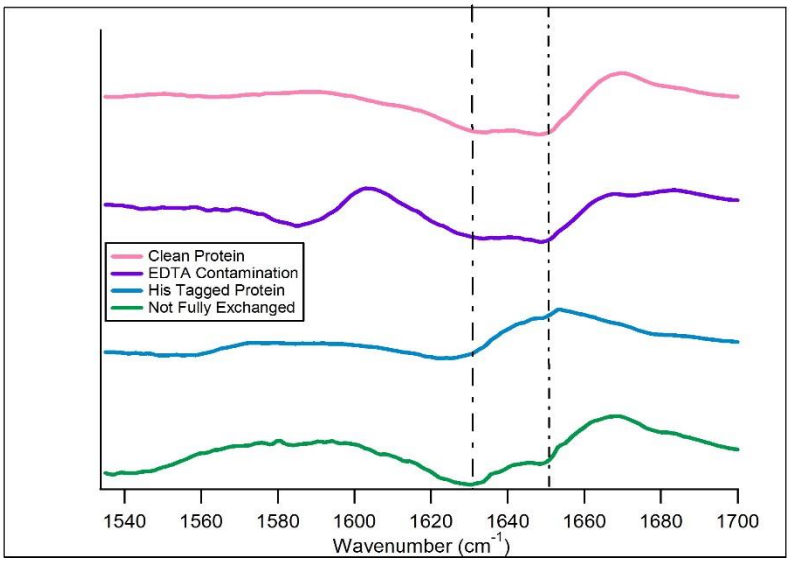


Figure 12: *FTIR spectra of samples taken throughout characterization process. Green curve shows spectrum for protein before optimization of exchange protocol. Blue curve shows effect of histidine affinity tag. Purple curve shows contamination of EDTA after development of cleavage protocol. Pink curve shows protein after removal of all anomalies and contaminates. Two dashed lines show 1632 cm^{-1} and 1648 cm^{-1} peaks for comparison across the spectra.*

The first development in the IR characterization of the BdpA proteins was to analyze a large broad peak around 1545 cm^{-1}. This peak corresponds to a structured peak that is observed when the nitrogen of the backbone residues is protonated. Since the heavy isotope label appears in this region, the peak is shifted by deuterating the bond. The protein is therefore exchanged in D_2O prior to taking infrared measurements. Any remaining protonated nitrogens will exchange with the deuterated solvent. This results in a temperature dependent peak that can obscure peaks of interest in the difference spectra. Exchange of the protein for a few hours was not sufficient to completely deuterate the protein. Therefore, the protocol for exchange of the protein was altered to increase both the temperature and the time of the exchange from room temperature for a couple of hours to 50°C overnight. Before the exchange parameters were adjusted, the IR difference spectra would show a feature similar to that seen in the green curve in Figure 12 around 1580 cm^{-1} thereby affecting the curve shape of the folded structure feature at 1632 cm^{-1}. By increasing the exchange time and temperature this feature was removed from subsequent IR spectra.

Another issue that arose in the characterization of the proteins by IR involved using the protein with the histidine affinity tag attached. Removal of the affinity tag from the protein involves costly reagents and greatly decreases the total protein yield. Therefore we conducted experiments to test the feasibility of using the affinity tagged version of the proteins in order to avoid the enzymatic cleavage. However, the tagged proteins showed a strong positive feature around 1650 cm^{-1} that would grow in until around 65°C and then vanish from the spectrum. This can be observed in the blue curve in Figure 12. This feature obscured the shoulder at 1648 cm^{-1} that also corresponds to a folded feature of the protein. After it was determined that the tag did affect the infrared signature of the protein, the previously described cleavage protocol was

developed to remove the tag from the protein. Subsequent IR spectra did not show the anomalous feature at low to mid-range temperatures.

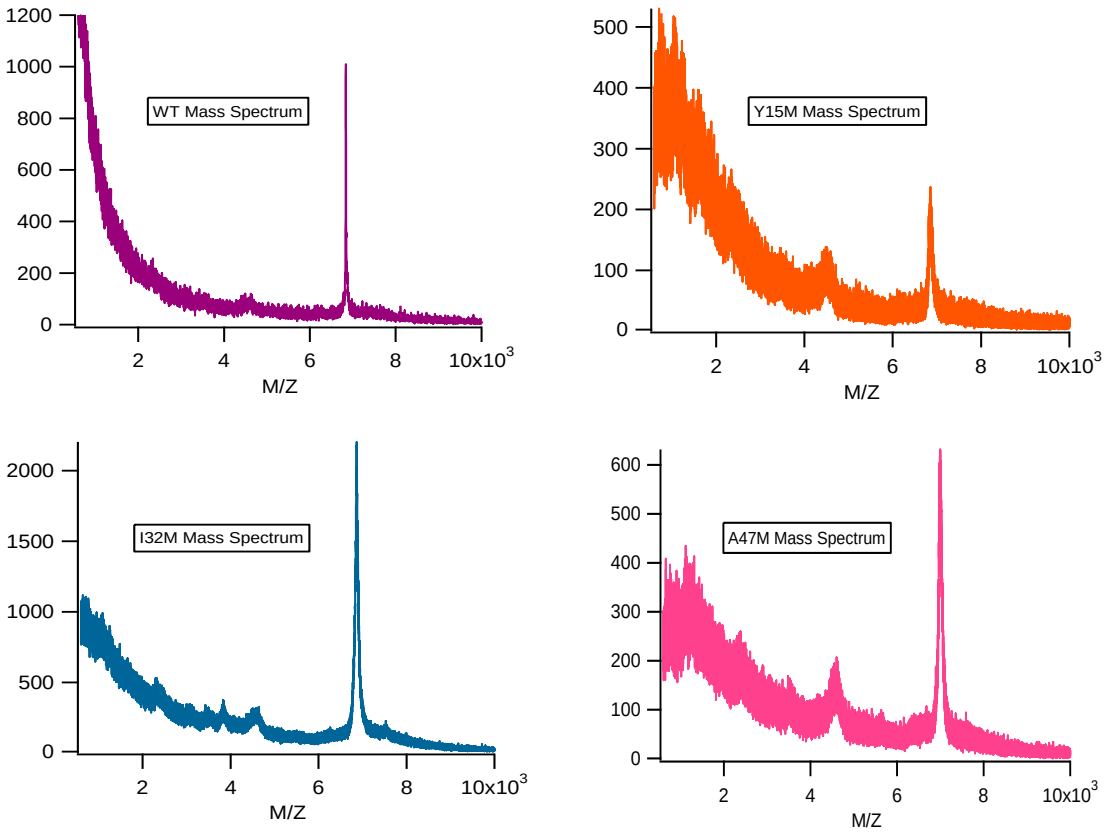
After the cleavage of the protein, the infrared spectra showed a peak around 1600 cm^{-1}. The intensity of this peak was typically near the intensity of the protein absorbance itself. The temperature change gave a derivative shaped feature that altered the baseline at the absorbance of the folded feature at 1632 cm^{-1}. Peaks that appear in the region of this contamination peak are typically carboxylic acids. It was therefore determined that the peak corresponded to EDTA that was bound to the protein from the cleavage buffer mixture. The peak assignment was confirmed by comparison to published data on the IR spectrum of EDTA in D_2O.[54] The previously mentioned $NiCl_2$ and HPLC purification techniques were developed to remove the EDTA from the samples. The final protein spectrum, seen in Figure 12 as the pink curve, shows none of the above mentioned anomalies and impurities and matches well with the published spectrum for wild type BdpA.[34]

Results and Discussion

In studying protein folding, it is essential to minimize perturbation to the native folding pathway when the measurement is made. Infrared spectroscopy is one of a few techniques that is able to achieve this by analyzing protein folding through heat denaturation without the addition of chemical denaturants that might perturb the pathway. However, detection of the folding pathway by infrared spectroscopy requires an infrared active label. Insertion of a label can also cause differences to the protein structure, thereby altering the folding pathway. Therefore, it is important to minimize this perturbation by introducing small labels that do not have an effect on the structure. This was accomplished by introducing a heavy isotope label that is the same

structure as the naturally occurring isotope. In the described protocol we have successfully introduced this heavy isotope label into three different mutant forms of BdpA in order to observe the thermodynamics and kinetics of the three different helices. A protocol for the expression of wild type BdpA was developed using an LB rich broth solution. A protocol for isolation using FPLC was established and the expressed protein was characterized by SDS-PAGE gel and mass spectrometry. The wild type protein gives a reliable yield of protein of around 20 mg of protein per liter of cell culture and the cells are stable on LB/agarose plates at 4°C. Glycerol stocks of the protein are also stable when kept at -80°C. A cleavage protocol was developed to remove the affinity tag from the protein and further purification was performed to remove additional contaminants. The protein itself is stable for several days in powder form after removal of any salts when kept in a desiccator for storage. Long term storage of the protein requires storage of the protein at -20°C. Once the wild type protein had been successfully expressed, a minimal media expression system was developed to express the mutants. The methionine mutants were first expressed in a solution including natural abundance methionine to develop the expression system before using the labeled methionine. After the proteins were successfully expressed with unlabeled methionine, the labeled methionine was introduced during the induction phase. The labeled mutants were isolated using the same isolation protocol as that developed for the wild type protein. The successful exchange of methionine was shown by mass spectrometry and infrared spectroscopy before use in the expression of the labeled mutants. The proteins were purified in the same manner as the wild type and the cleaved proteins were obtained and purified. The mutants can be stored in the same manner as the wild type. The mass spectra for the cleaved version of all four proteins can be seen in Figure 13 showing the successful isolation of the proteins.

Infrared spectroscopy was used as a characterization method throughout the development of the expression protocol for the proteins. Several problems were identified and addressed through various improvements in the protocol that was originally developed. The infrared spectrum of the pure protein shows the two expected features of the folded protein at 1632 cm^{-1} and 1648 cm^{-1} and the broad unstructured feature at higher temperatures around 1670 cm^{-1}. The region where the label appears around 1560 cm^{-1} is transparent in the wild type protein spectrum and the feature can be readily observed in each of the mutant proteins. The established protocol allows for the expression of BdpA and the three developed mutant forms with useful infrared characteristics for further analysis.



Mass (Da)	WT	Y15M	I32M	A47M
With Tag Expected	8576	8544	8594	8636
Cleaved Expected	6827	6798	6849	6891
From Mass Spec	6837.3	6852.7	6864.1	6992.5
Salts Present	-	$2Na^+$	-	$3Na^+$ $1Cl^-$

Figure 13: *Mass spectra of all four cleaved proteins. Mass spectra of mutants are for labeled mutants.*

Table 3: *Tabulated data of masses of cleaved and uncleaved proteins and masses gathered from mass spectra. Masses that differ from predicted mass of the protein correspond to protein in the presence of salt. Salts present in samples are shown on the final row.*

Chapter 2 Thermodynamics

Introduction

Secondary structural features can be used to determine the thermodynamic stability of a protein by following the temperature dependence. Two techniques commonly used to measure these temperature dependent changes are circular dichroism (CD) spectroscopy and infrared (IR) spectroscopy. CD spectroscopy monitors changes in the differential absorption of left and right circularly polarized light of the protein as temperature is increased. By monitoring this value for a specific wavelength corresponding to a secondary structural feature, a melt temperature can be determined for the protein. The melt temperature is defined as the temperature at which 50% of the protein is in the folded state and 50% in the unfolded state for a cooperative two state transition. This value gives a measure of the stability of the protein. A corresponding technique involves taking equilibrium Fourier transform infrared (FTIR) measurements over a range of temperatures. As the protein is denatured by heat, peaks corresponding to the folded structure will lose intensity and peaks corresponding to the unfolded structure will gain intensity. These changes in absorbance intensity can be monitored over a range of temperatures. If a protein contains several secondary structure features, such as both helices and sheets, then different absorbances will be observed corresponding to the frequency of the vibration of the backbone residues in that feature.

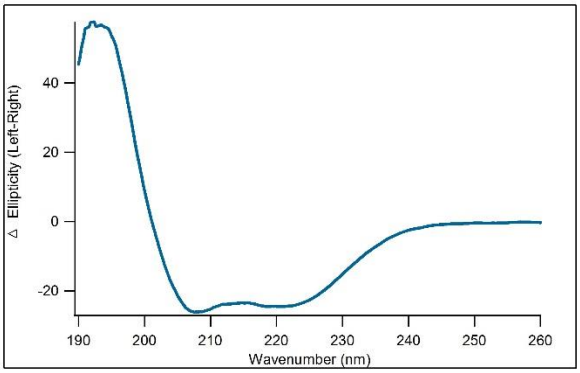


Figure 14: *Characteristic CD spectrum for alpha helical proteins*

Circular dichroism (CD) spectroscopy can be used to monitor changes in secondary structure over a range of temperatures. Secondary structure features also give characteristic absorbances

of circularly polarized light in the ultraviolet region. A typical spectrum for alpha helical proteins can be seen in Figure 14. For alpha helical proteins, there are two minima at around 222 nm and 209 nm. The minimum at 222 nm is typically chosen for observation of alpha helix content in a protein.[56] This peak is then monitored over a range of temperatures in order to determine the melt temperature of alpha helical proteins such as BdpA.

A similar absorption study can be conducted using infrared spectroscopy. The largest feature that is generally used in IR observation is the amide I band (amide I' in deuterated solvents) and corresponds to the stretching frequency of the carbonyl bond in the backbone of each amino acid. In alpha helical proteins, two absorbances can be seen. The first is for solvated helices and appears around 1632 cm^{-1}, the second corresponds to buried helices appear around 1648 cm^{-1}.[55] The structure of the backbone and the corresponding alpha helical absorbances can be seen in Figure 15.

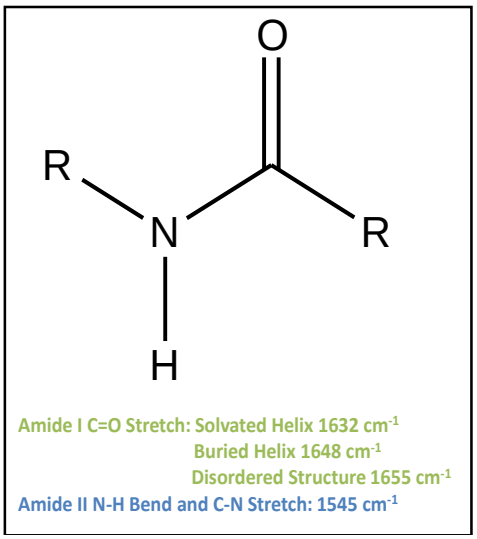


Figure 15: *Structure of backbone of protein showing carbonyl. Vibrations that correspond to the various monitored infrared wavelengths are listed below the structure.*

Site specific infrared peaks can be observed by the introduction of a label that gives an absorbance in a transparent region of the protein's infrared spectrum. A label that can be introduced with no perturbation to the protein structure is a heavy isotope version of the carbonyl that gives rise to the amide I peak. An ideal heavy isotope is a $^{13}C=^{18}O$ carbonyl bond in the backbone at a single position. Based on the harmonic oscillator approximation, an infrared vibrational frequency will be inversely related to the reduced mass of the atoms on either side of the oscillator, in this case the bond. By using a double heavy isotope label, the observed peak will shift even further from

the corresponding unlabeled peak than a single heavy isotope label making it easier to resolve. This approximation relies on the oscillator being independent from coupling with any neighboring oscillators. Coupling would cause a splitting of the signal which is not observed in the label peak of the difference spectrum.[56] Therefore, the reduced masses can be used to determine the expected shift in the wavelength. When the reduced masses for the heavy isotope label and natural abundance isotope oscillators are used in the equation, the estimated shift in the wavelength is around 75 cm^{-1}.[48] The shift of the heavy isotope label peak corresponds to around 70-72 cm^{-1} from the 1632 cm^{-1} solvent exposed unlabeled peak in the spectra, agreeing well with the expected shift from the approximation. This shift shows that a unique signal for a given secondary structure feature will be introduced by incorporating a heavy isotope label at that position. Therefore, a label can be introduced at a specific site in the protein sequence and used to monitor the thermodynamic properties of the protein at that site.

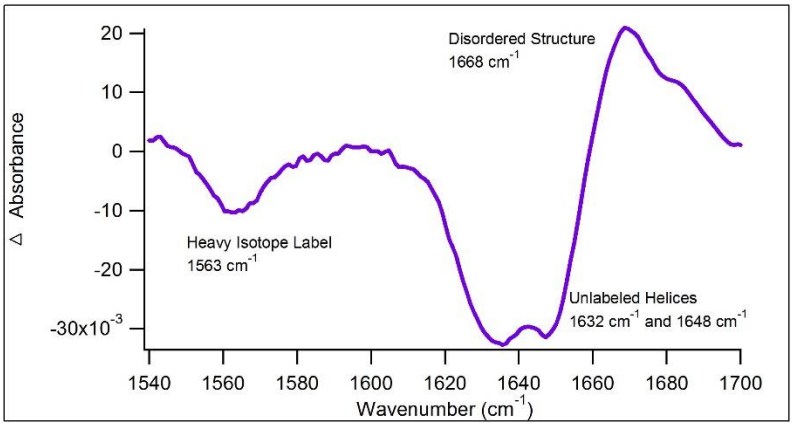


Figure 16: *Spectrum showing heavy isotope shift in vibrational frequency of backbone carbonyl of alpha helix.*

The change in absorbance as the protein folds or unfolds can be directly observed and correlated to the concentration of the folded species for a given temperature. The stability of a protein can be analyzed by obtaining a temperature dependent melt curve for the observed wavelength. The protein stability is a good measure of the level of perturbation introduced by creating mutant versions of a protein. Ideally, the melt temperatures will be within error of one another, but a lower melt temperature will indicate that the structure of the protein has been altered

enough by the mutation to cause a change to the structure and potentially the folding pathway of the protein. Since the heavy isotope label is introduced by the previously described point mutation procedure, it is important to be able to compare the stabilities of each of the mutant proteins to that of the wild type.

For circular dichroism, the melt curve is constructed by plotting the delta ellipticity over a range of temperatures. The corresponding curve can be fit using an appropriate fit function to determine a melt temperature (T_M). The melt temperature will be higher for more stable proteins. The melt curve also shows what percentage of the protein is folded at each temperature.

To construct an FTIR melt curve, IR spectra are collected for the sample solution and a reference solution over a range of temperatures. The reference spectra can be used to account for any absorbance features that correspond to the solvent, typically D_2O, or other components in the buffer solution. Difference spectra are then calculated by subtracting the lowest temperature spectrum from each of the higher temperature spectra. These difference spectra are plotted to determine the induced absorbance or bleach of absorbance at a given wavelength with changing temperature. The intensity at the wavelength of interest is plotted against temperature to construct a melt curve. The melt curves are then fit using an appropriate fit function to determine the melt temperature from the infrared spectra. The determined melt temperatures calculated from CD spectroscopy and IR spectroscopy should be comparable if the same secondary structure feature is being probed.

In the infrared spectra the location of the label peak can be identified for use in collecting the kinetic data. The nonhomogenous distribution of structures in the disordered structure of the label causes the corresponding induced absorbance to be indiscernible from the baseline;

however, the bleach of the structured peak can be clearly seen and monitored over changing temperatures. The derived curves will give the melt temperature for each of the proteins and a comparison of the stability of the label position to the overall protein stability. By determining the stability of each of the proteins, the level of perturbation introduced by the labels can be determined.

Materials and Methods

CD Spectroscopy

CD spectra were collected using a JASCO J-810 with a peltier temperature controller (JASCO, Easton, MD). A background spectrum of water was collected in order to assure that the noise was within 1 mdeg. A spectrum was then measured of the protein solution from 190 nm to 290 nm. The protein solutions were prepared by dissolving the protein in deionized water to a concentration in the range of 100's of μg/mL. CD melt curves were collected by measuring the delta ellipticity over a range of temperatures from 20°C to 105°C, collecting a data point every 1°C. The spectrometer was set up to scan at a rate of 100 nm/min with a response of 2 seconds and a band width of 4 nm. The temperature was ramped at 40°C per hour. The collected curves were then fit to determine a melt temperature. After collection of the melt curve another spectrum was taken to determine if the protein was able to fully recover the degree of helicity of the first spectrum.

Protein and Cell Preparation

For infrared spectroscopy, dried proteins were exchanged overnight at 50°C in D_2O. The solvent was then removed by lyophilization. The dry exchanged protein was dissolved in a phosphate

buffer to solutions with a concentration around 5 mg/mL. The phosphate buffer used was composed of 25 mM potassium phosphate and 50 mM NaCl at a pH of 6.75. Buffer was lyophilized and deuterated before dissolving protein samples.

For taking infrared measurements, samples were injected into a custom built cell. The cell contains two CaF_2 windows that are transparent in the infrared region. A 130 µm spacer between the two windows creates two compartments for loading the reference buffer solution and the protein solution on either side. The windows are held in a copper cell casing that can conduct heat to the protein sample.

IR Spectroscopy

Equilibrium Fourier transform infrared spectroscopy was used to evaluate the infrared spectrum of each sample. Spectra were collected on an Excalibur 3100 instrument from Varian (Agilent Technologies, Santa Clara, CA) with a speed of 40 kHz at a resolution of 2 cm^{-1}. The detector used was a mercury cadmium telluride detector cooled by liquid nitrogen. The instrumental setup was purged using dehumidified air before scanning. The aperture was adjusted so that the centerburst voltage was approximately 5 volts. Scans were collected over a range of 400-4000 cm^{-1} for 256 scans that were then averaged using an in house developed Lab View program (National Instruments, Austin, TX). Spectra were collected for the reference sample and for the protein sample at each measured temperature. Spectra were collected every 5°C from 20°C to 105°C as determined by the temperature of the water bath. A thermocouple was attached to the copper cell to record the accurate temperature of the protein solution. The data were evaluated using Varian Resolutions Pro version 4.1.0.101 (Agilent Technologies, Santa Clara, CA).

Infrared Data Evaluation

Raw transmission spectra of the sample were ratioed to the reference transmission spectrum to get the raw absorbance data of the protein. Changes over temperature can be seen by comparing the raw absorbances at each temperature to determine where signal is lost or gained. However, since the total signal change from the lowest to the highest temperature is generally an order of magnitude smaller than the total absorbance, it is more useful to examine difference spectra in order to determine the changes that occur with temperature. Difference spectra are constructed by subtracting the lowest temperature spectrum from each of the higher temperature spectra. Peaks that correlate to unfolded structure will increase in intensity and show an induced absorption as the temperature is increased. Peaks correlating to folded structure will decrease in intensity and show a bleach in absorption as the temperature is increased. The intensity of the absorbance of a specific wavelength at each temperature can be plotted to construct a melt curve. Fitting the curve will give the melt temperature of the protein at this wavelength.

Fit Data

The derived melt curves can be fit by a sigmoid if the baseline does not change with temperature. This implies that the transition is a two-state transition from a folded state to an unfolded state without any thermodynamic intermediates. However, shifts in the solvation of the protein can cause changes in the baseline of the curve. Therefore, the curves were fit with a single or double baseline sigmoid fit function using IGOR Pro 5.0 (WaveMetrics, Lake Oswego, Oregon) as shown below.

Single baseline fit: $$f(x) = m*x + z + \frac{max}{1+e^{\frac{xhalf - x}{rate}}}$$

Multiple baseline fit: $$f(x)=1+A*x+B+\frac{(C*x+D)*e^{\frac{xhalf-x}{rate}}}{1+e^{\frac{xhalf-x}{rate}}}$$

By accounting for the sloping baselines, the sigmoidal region can be accurately fit and a melt temperature determined for the protein. These fit curves were employed for the melt curves derived from the CD spectra and the IR spectra depending on whether the baselines were sloped.

Calculation of Free Energy

Free energy values were calculated for 25C for each of the samples using the following equation:

$$\Delta G_f = -RT \ln K_f$$

Free energy values were used to determine if the protein structure had been significantly altered by introducing the mutations.

Results and Discussion

When constructing mutant versions of proteins it is essential to determine if there has been any perturbation to the protein structure in order to ensure that the protein will not lose its functionality, its folded structure, or have an alternative folding pathway. In monitoring protein folding, CD and IR spectroscopy are used to measure melt temperatures as a measure of stability. The melt temperature was determined using both spectroscopy methods for all four of the proteins, the wild type protein and each of the mutants. Furthermore, infrared spectroscopy

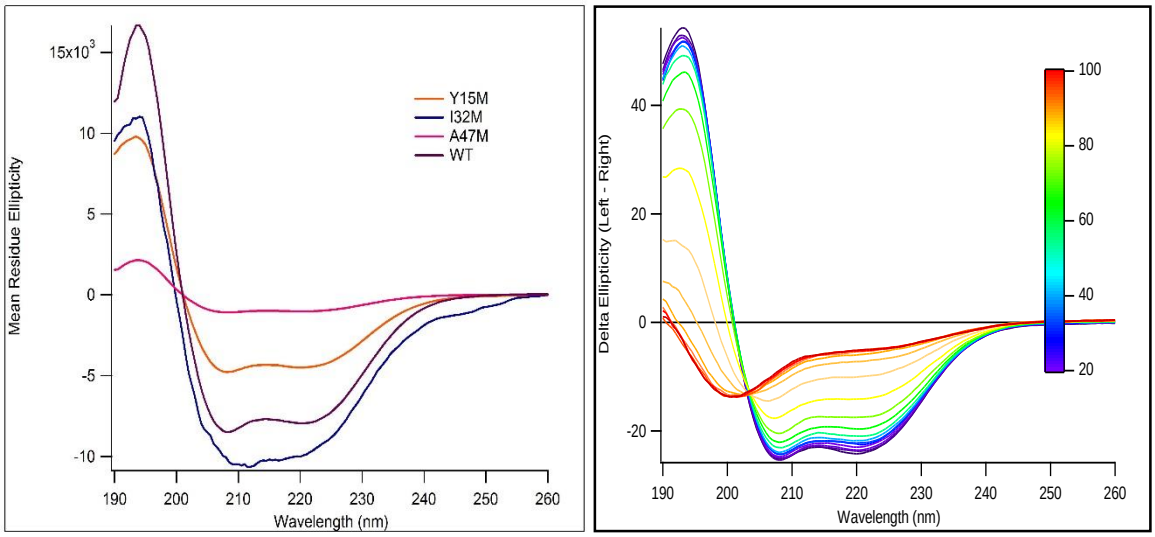


Figure 17: *Left – Initial CD spectra collected for each protein.*

Right – Sample of spectra collected at each temperature during temperature melt.

was also able to show what peaks were present in the wild type protein and the additional peak present in the labeled mutant indicating the wavelength for the labeled carbonyl.

The first spectroscopy used was circular dichroism. The CD spectrum for each of the proteins can be seen in Figure 17. The spectra were corrected for differences in concentration by calculating the mean residue ellipticity for each wavelength. Each of the proteins showed a high

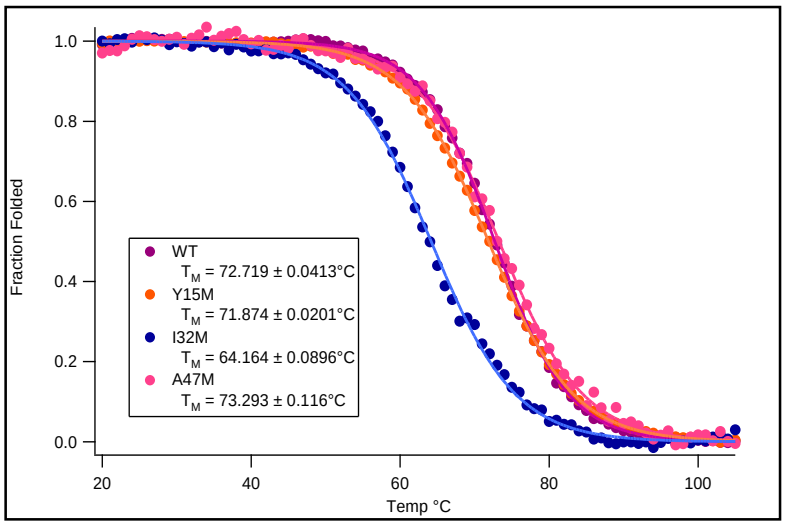


Figure 18: *Melt curves for each protein derived from CD spectroscopy. Determined T_M values are given for each protein.*

degree of helicity as seen in the shape of the collected spectra, indicating that the amount of helical structure was not altered by the mutations in the mutant proteins as compared to the wild type. Differences in the mean residue ellipticity of the different proteins for the various wavelengths indicate there might be a small difference in helical content between the proteins; however, slight errors in the calculation of the sample concentration could also be the cause. Melt curves were determined for each of the proteins by monitoring the change in intensity of the minimum at 222 nm. Figure 17 shows the spectra collected at each temperature as the temperature was increased. The derived melt curves were normalized to adjust for contributions from solvent reorganization and determine the melt temperature for each of the proteins. The derived melt curves and corresponding melt temperatures are shown in Figure 18. The melt temperatures indicate that the helix one and helix three mutants (Y15M and A47M respectively) have very similar stabilities to the wild type protein. The only mutant to show a distinct difference is the helix 2 mutant (I32M). The lower stability of this mutant indicates that the mutated amino acid plays a more significant role in the proper formation of the protein structure. The similar slopes of the sigmoid regions show that the cooperativity of the protein folding is the same for these proteins. A highly cooperative transition occurs when the entire structure of the protein unfolds simultaneously. This will give a very steep sigmoidal melt curve since the unfolding occurs as one rapid step. Therefore, the cooperativity can be compared between normalized curves to determine if the folding has been significantly altered in terms of how cooperative the transition is.

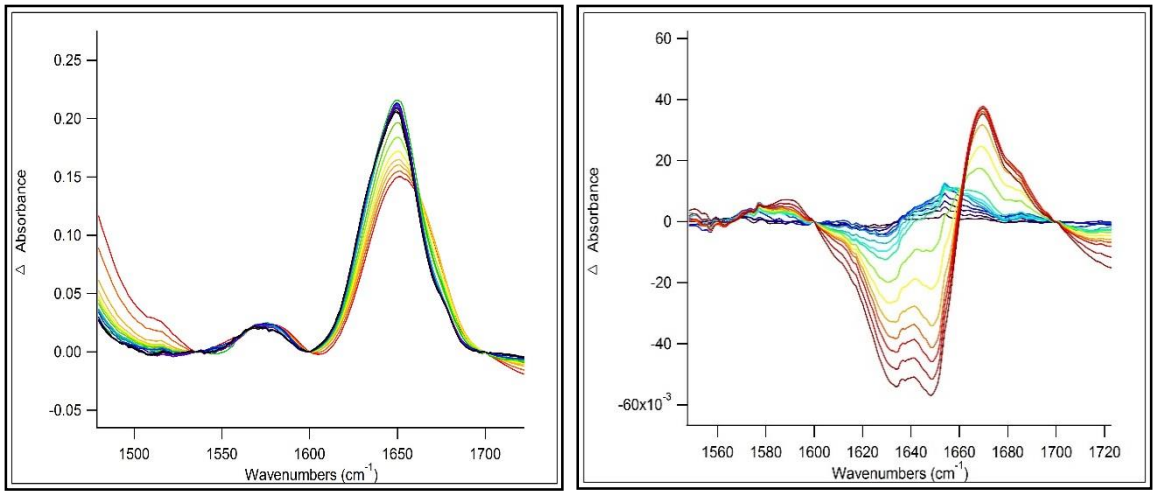


Figure 19: *Left – Absorbance spectra of wild type.*

Right – Difference curves constructed for wild type.

After the CD data was collected, the protein stability and cooperativity were also studied using FTIR spectroscopy. Figure 19 shows a sample absorbance series over change in temperature for wild type BdpA and the difference curves derived from the absorbance spectra. Figure 20 shows a sample absorbance series for the helix 1 mutant, Y15M, and the derived difference spectra. Melt curves were constructed for the peaks corresponding to the solvated helix, the buried helix, and in the mutants the labeled position for each of the proteins. The curves were also

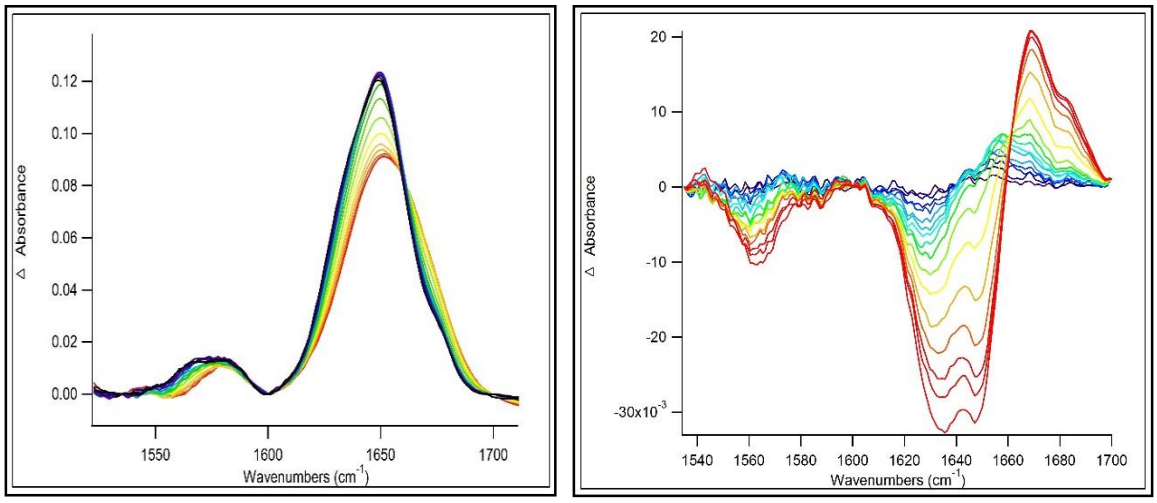


Figure 20: *Left – Absorbance spectra of Y15M to show mutant peaks.*

Right – Difference curves constructed for Y15M to show mutant peaks.

normalized to adjust for solvent contributions. The normalized melt curves for each of the monitored peaks are shown in Figure 21. The curves from the 1648 cm^{-1} absorbance, which corresponds to the buried helix, show little change in the baselines before and after the melt temperature even before normalization. This is due to the fact that solvent will have little effect on the buried residues since they are protected from the solvent with the interior of the protein. The curves from the 1632 cm^{-1} absorbance, however, show significantly sloping baselines before normalization due to interaction with the solvent. Therefore, the sigmoidal shape of the 1648 cm^{-1} melt curves results in the derivation of more reliable melt temperatures since baseline corrections are minimal. In the labeled peak melt curves the signal level is much smaller than for the other protein absorbances introducing a high level of noise. This can be seen in the normalized curves and in the error in the fits of the melt curves in Table 4. The label peak melt curves do show that the label has a similar temperature dependence as compared to the unlabeled absorbances and this dependence can be used to monitor the kinetics. Once again, the melt temperatures and cooperativity are similar for all of the proteins excluding the helix 2 mutant (I32M). This protein shows a lower melt temperature in the infrared melt curves as well as the CD, supporting the hypothesis that the introduction of the mutation caused a change in stability. However, this change in stability does not significantly impact the thermodynamic properties of the protein as can be seen by comparing the protein free energy values. The free

	CD 222 nm kcal/mol	1560 cm^{-1} kcal/mol	1632 cm^{-1} kcal/mol	1648 cm^{-1} kcal/mol
WT	-2.8 ± 1.5	--	-3.4 ± 1.5	-3.3 ± 1.5
Y15M	-3.0 ± 1.5	-2.1 ± 1.5	-1.8 ± 1.5	-3.4 ± 1.5
I32M	-2.9 ± 1.5	-1.2 ± 1.5	-3.5 ± 1.5	-2.9 ± 1.5
A47M	-2.0 ± 1.5	-1.7 ± 1.5	-2.8 ± 1.5	-2.8 ± 1.5

Table 4: *Table of free energy values calculated for CD and IR spectroscopy at 25°C*

Sample	CD	IR		
	222 nm	1563 cm^{-1}	1632 cm^{-1}	1648 cm^{-1}
WT	72.6 ± 0.7°C	--	72.0 ± 0.7°C	73.6 ± 0.7°C
Y15M	71.9 ± 0.7°C	67.2 ± 2.2°C	67.9 ± 0.7°C	70.7 ± 0.7°C
I32M	64.2 ± 0.7°C	48.3 ± 7.1°C	64.6 ± 0.7°C	65.5 ± 0.7°C
A47M	73.3 ± 0.7°C	67.5 ± 1.1°C	73.7 ± 0.7°C	74.5 ± 0.7°C

Table 5: *Table of T_Ms derived from melt curves from CD and IR spectroscopy*

energy calculations for each of the proteins show that no significant change to the protein was caused by any of the mutations introduced and as can be seen in Table 4.

The degree of helicity was shown to be unaffected by introducing mutations through the use of CD spectroscopy. The infrared spectra also showed that the mutants have similar secondary structures to the wild type protein. The melt temperatures determined through both spectroscopies indicate that the stability of the protein was not significantly altered by adding the mutations and are tabulated in Table 5 for comparison. The only protein to show a slight decrease in stability as compared to wild type was the helix 2 mutant (I32M). The lower stability of the helix 2 mutant indicates that the residue where the mutation was inserted could be important in the folding pathway of the protein. However, the free energy of the protein was within error of the free energy of the wild type showing that the destabilization did not have a significant effect on the protein's thermodynamic properties and the mutant is viable for kinetic studies. The peaks that were used in the derivation of the melt curves can be monitored to determine the kinetics for each of the variations on the backbone signal. The absence of additional peaks in the infrared spectra indicate that the measured kinetics will be solely derived from changes to the protein structure and can therefore be used in the interpretation of a folding pathway for the protein.

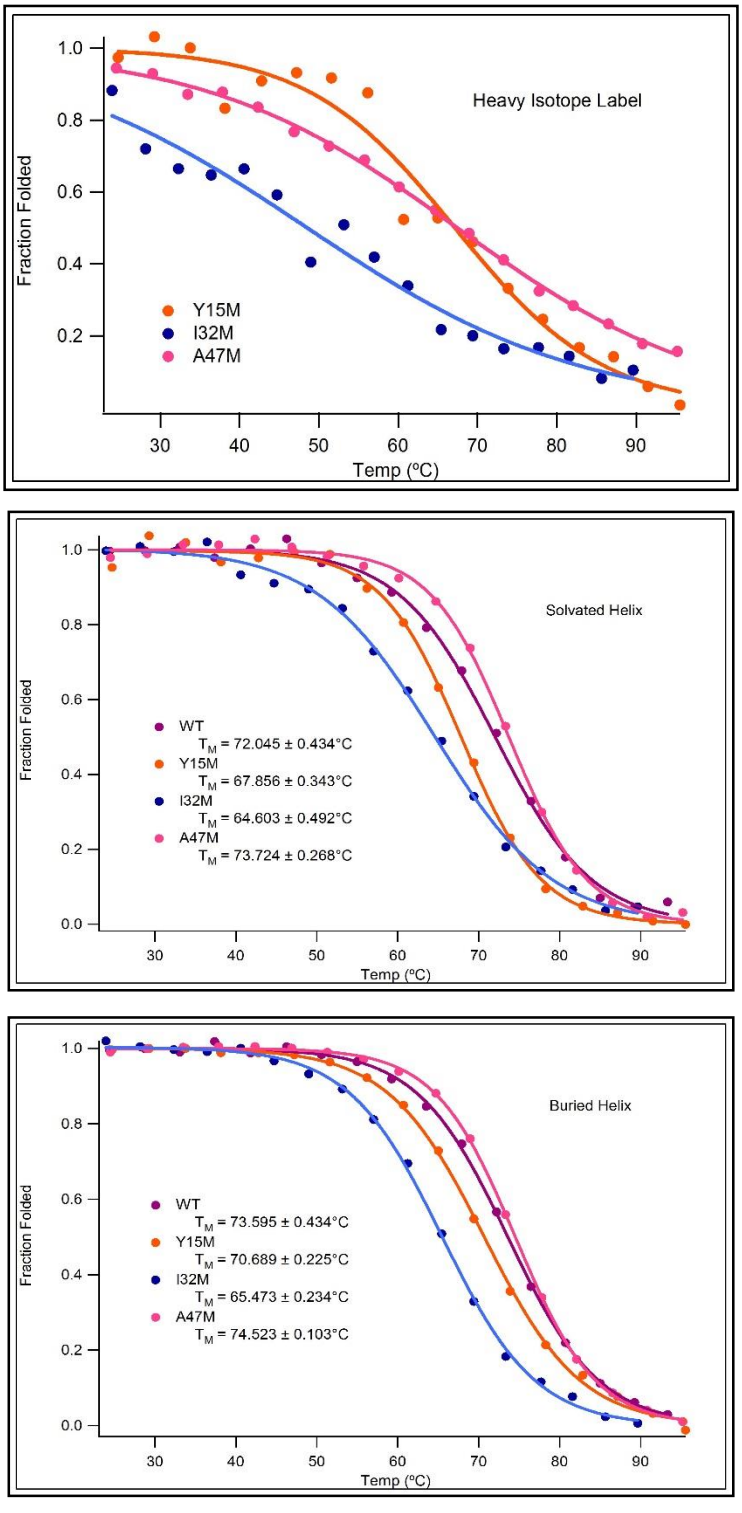


Figure 21: *Top – Melt curves for labeled peak at 1560 cm^{-1}.*

Middle – Melt curves for solvated helix at 1632 cm^{-1}.

Bottom – Melt curves for buried helix at 1648 cm^{-1}.

Chapter 3 Kinetics

Introduction

Infrared spectroscopy is also a useful tool in determining the folding pathway of a protein through the determination of the kinetic events that take place as the protein folds. For ultrafast proteins, the kinetic events of interest occur within nanoseconds to microseconds. This requires an infrared spectrometer that is able to detect changes in absorbance on this fast of a timescale. Such time scales can be accessed using temperature-jump infrared (T-jump IR) spectroscopy.[57,58] T-jump spectroscopy has been used for studying relaxation rates of proteins since the method was pioneered by Dyer et al in 1996.[59] T-jump IR spectroscopy uses a pulsed laser system to briefly and repeatedly heat a protein system, perturbing the equilibrium to a new equilibrium with a higher proportion of unfolded sample as seen in Figure 22. Even heating of the sample can be achieved within 10 ns.

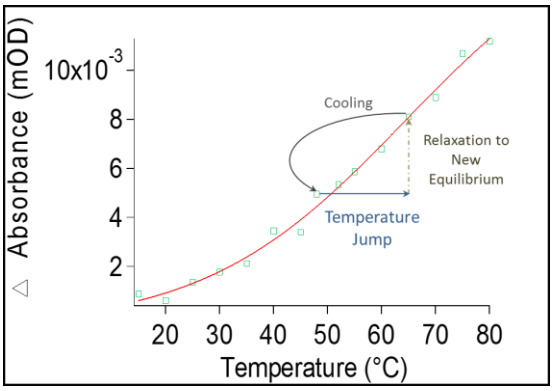


Figure 22: *Spectrum showing relaxation to equilibrium induced by temperature-jump and how it corresponds to the absorption changes as seen in a melt curve.*

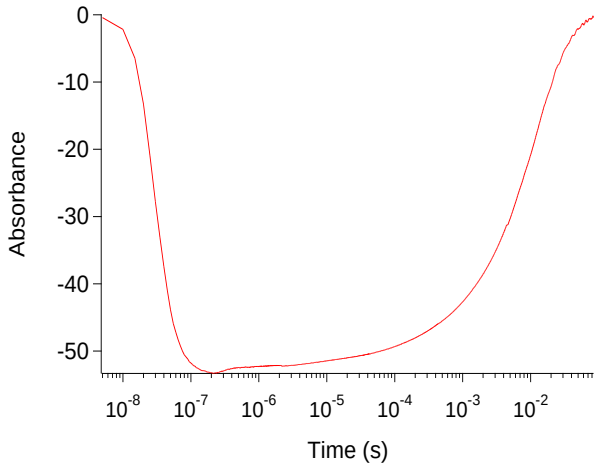


Figure 23: *Sample spectrum collected with a temperature-jump*

As the sample returns to its original temperature, the equilibrium is also reestablished with the additional unfolded population relaxing back to its folded state. During this process, absorbance at a specific wavelength is determined by changes in the detected signal of a

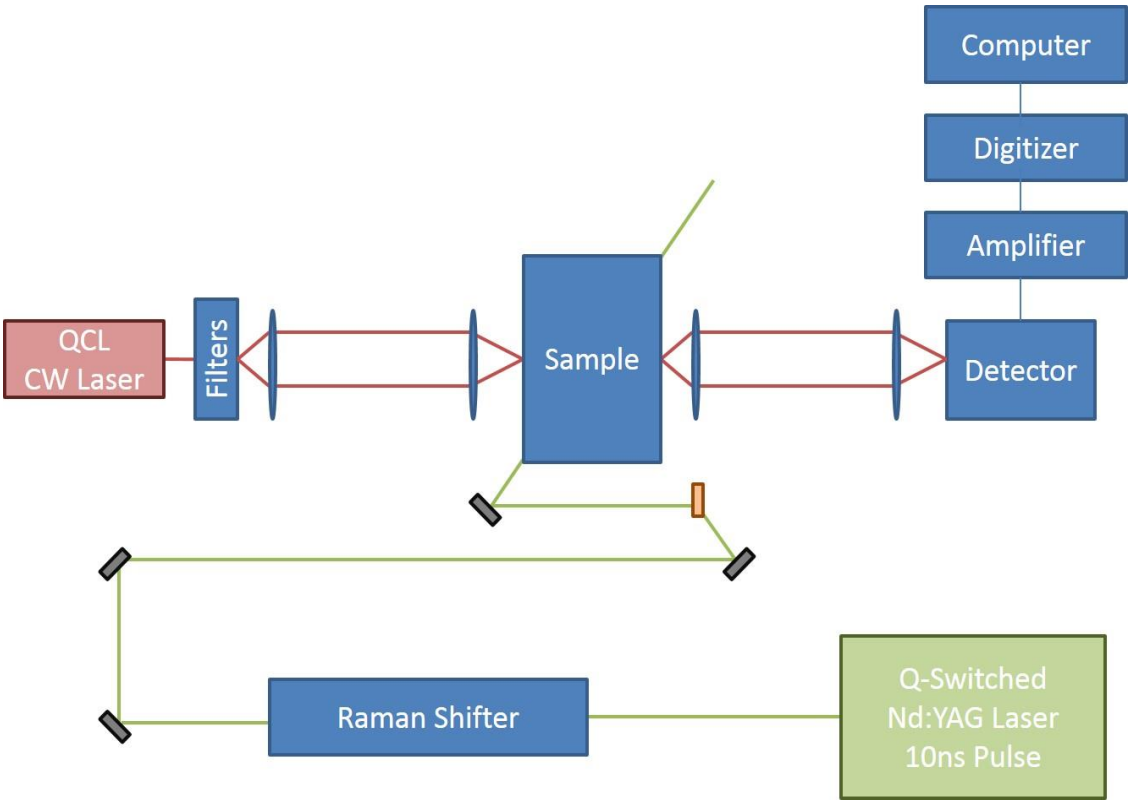


Figure 24: *Schematic of a nanosecond laser infrared T-jump spectrometer.*

second laser tuned to the wavelength of interest. An example of a typical T-jump setup is shown in Figure 24. The signal initially experiences a large change in absorption due to the temperature dependence of the D_2O present in the solvent.[60] The protein system then relaxes to equilibrium at the higher temperature of the solvent on the timescale of nanoseconds to milliseconds. A second large change in absorption is then observed as the system cools back to the initial temperature. The measured change in absorbance over time can be deconvolved, resulting in time constants for any kinetic phases that occurred. The temperature range selected for the jump is chosen to be near the T_M of the protein since this is where the greatest change in signal will be observed due to the sigmoidal nature of protein melt curves. Background noise and absorption caused by the buffer solution are also corrected for before calculating the time constants by collecting data at a reference wavelength known to be of zero absorbance across a

range of temperatures for the protein. The resulting deconvolution can therefore be assumed to give time constants specific to the protein itself.

What makes this form of spectroscopy especially powerful is the ability to monitor kinetic rates at a specific wavelength. By monitoring several wavelengths, the kinetic rates for different structural features can be determined. Since a labeled bond will absorb at a different wavelength than an unlabeled bond, the label kinetics can be determined separately from the kinetics of the protein. This allows for the deterimining highly localized kinetics corresponding to when a specific portion of the protein folds. Monitoring these rates for labels placed in various sites of a protein can give data for each localized region of the protein which, when compiled, will give a clear picture of how the protein folds.

Materials and Methods

Protein and Cell Preparation

For infrared spectroscopy, dried proteins were exchanged overnight at 50°C in D_2O. The solvent was then removed by lyophilization. The dry exchanged protein was dissolved in a phosphate buffer to solutions with a concentration around 5 mg/mL. The phosphate buffer used was composed of 25 mM potassium phosphate and 50 mM NaCl at a pH of 6.75. Buffer was lyophilized and deuterated before dissolving protein samples.

For taking infrared measurements, samples were injected into a custom built cell. The cell contains two CaF_2 windows that are transparent in the infrared region. A 130 µm spacer between the two windows creates two compartments for loading the reference buffer solution and the protein solution on either side. The windows are held in a copper cell casing that can conduct heat to the protein sample.

Temperature-Jump Spectroscopy

Temperature-jump experiments were conducted using a pump probe method with a setup similar to that pictured in Figure 24. The sample was mounted onto a stage that is temperature controlled by a water bath. A rapid temperature change (10 ns) was initiated by an Nd-YAG laser (Spectra Physics, Mountain View, CA) at 1064 nm. This pulse is Raman shifted using H_2 gas to a 2 µm beam that heats a small volume of the sample. The D_2O in the sample was heated by the beam, and the sample relaxed to the new equilibrium state at the higher temperature. The beam was set to a 10 Hz pulse in order to initiate the rapid temperature change repeatedly for averaging of the signal. The probe signal was provided by a tunable quantum cascade laser (Daylight Solutions, San Diego, CA) with a tunable range of 1525 cm^{-1} to 1690 cm^{-1}. The signal was detected by a mercury cadmium telluride detector cooled by liquid nitrogen. The signal is recorded from 10^{-9} to 10^{-2} s with a Tektronix 7912 digitizer (Tektronix, Beaverton, OR). A reference and sample signal were recorded for each wavelength.

Kinetic Data Evaluation

The reference and sample signals were scaled to one another and subtraction of the reference signal yielded the sample kinetic trace. The kinetic trace can be fit by a single or double exponential using IGOR Pro 5.0 (WaveMetrics, Lake Oswego, Oregon) depending on the number of kinetic phases present. The fit functions included an offset for both the x and y direction and the formulas used are shown below.

Single Exponential: $$f(x) = y_0 + A * e^{\left(\frac{-(x-x_0)}{\tau}\right)}$$

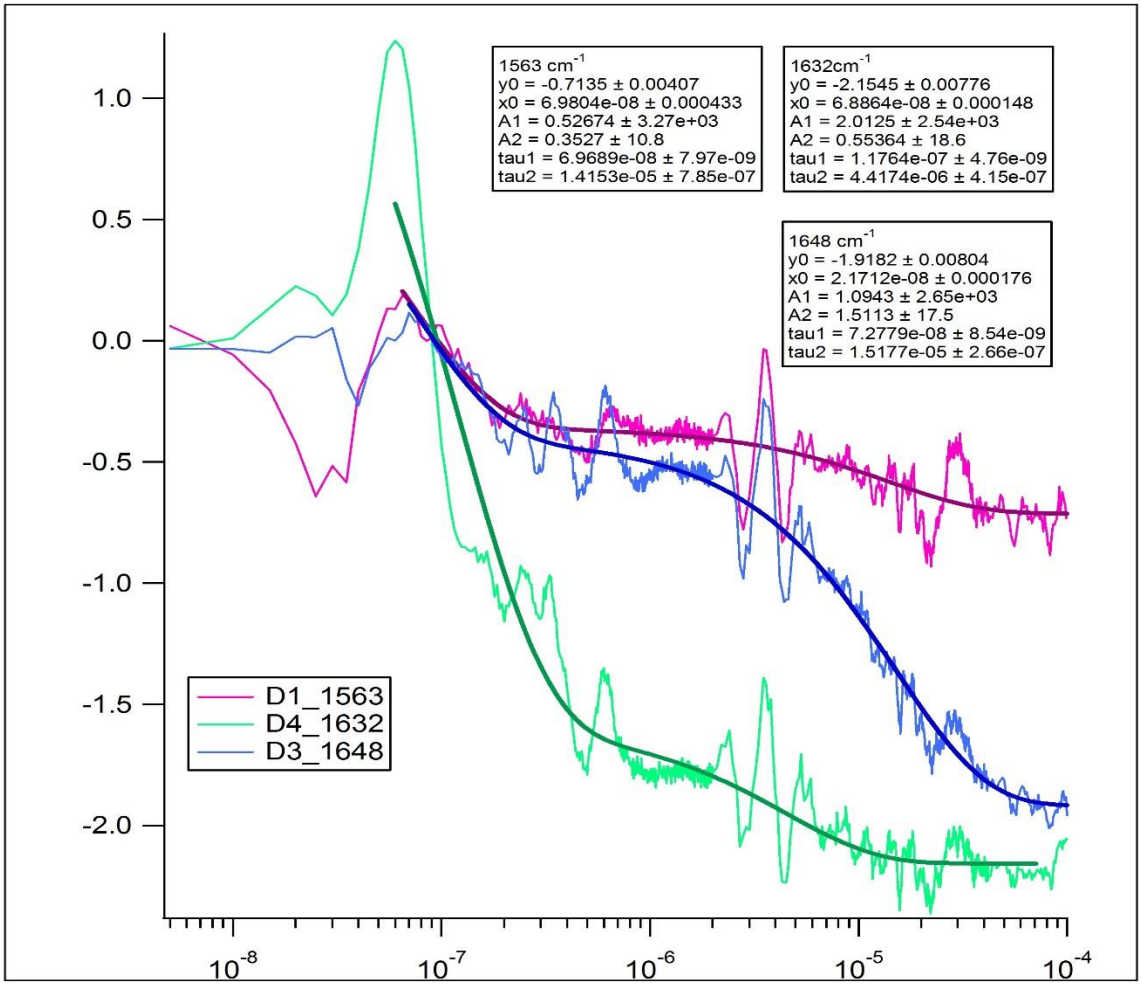
Double Exponential: $$f(x) = y_0 + A_1 * e^{\left(\frac{-(x-x_0)}{\tau_1}\right)} A_2 * e^{\left(\frac{-(x-x_0)}{\tau_2}\right)}$$

Results and Discussion

Preliminary kinetics data were gathered for the wild type protein and the helix three mutant. The kinetics of the wild type matched well with those previously reported assuring that the protein had been expressed and recovered properly.[34] The kinetics of the mutant show interesting preliminary results as seen in Figure 25. The time constants for both proteins observed and the published values for the wild type protein are shown in Table 6. The curves show a high frequency noise that is the result of several factors. Differences in the polarization between the reference buffer signal and the sample signal can cause artifacts in this region. Problems with the alignment of the QCL laser can also give rise to noise on this timescale. This noise can be inherent in the laser or the result of the alignment between the pump and probe lasers. These problems can be minimized by using a reference wavelength on the sample side instead of collecting a spectrum with a reference buffer since the polarization and the output from the laser will be much better matched than the signals between two different solutions even at the same wavelength.

The solvated helix wavelength shows a large contribution of a fast timescale on the order of 100 ns and a small contribution of a slower timescale on the order of 4 μs. These values and the relative contributions of each phase are similar to the reported and measured time constants of the wild type protein.[34] The faster microsecond phase observed for the mutant at this wavelength could be the result of a change in the kinetic properties of the protein or simply an issue with the resolution of the signal from the noise. The buried helix wavelength shows two phases as well, and similar contributions of the fast and slow phases when compared to the wild type. However, the fast kinetics phase is difficult to deconvolute from the response of the detector which is around 50 ns. The labeled peak offered an interesting evaluation by showing

two kinetics phases as well. The fast kinetics phase is close in scale to the phase expected for the unlabeled positions, suggesting that it is present in the intermediate state of the protein. However, it is also close in time to the response of the detector and may therefore be an artifact of the detector itself. By using a faster detector this phase can be resolved and conclusively assigned to either the label kinetics or the detector response. The slower kinetic phase is similar to that reported for the wild type protein and for the value determined for the unlabeled slow phases. This shows definitively that kinetic data can be determined for the label position by minimizing the noise from the system and using a faster detector to resolve the fast phase that was observed. The proteins can therefore be subjected to temperature-jump spectroscopy to determine a folding pathway for the protein. The labeling methodology that was developed will be of use in further study of this protein.



.

	1560 cm^{-1}		1632 cm^{-1}		1648 cm^{-1}	
	τ_1 (ns)	τ_2 (µs)	τ_1 (ns)	τ_2 (µs)	τ_1 (ns)	τ_2 (µs)
Reported	--	--	90 ± 10	9.7 ± 1	200 ± 50	8.6 ± 1
WT	--	--	140 ± 20	9.2 ± 1	130 ± 20	15 ± 4
A47M	70 ± 4	14 ± 6	120 ± 20	4.4 ± 1	73 ± 20	15 ± 4

Figure 25: *Collected kinetic data for helix three mutant (A47M) taken at 50°C with a temperature jump of 15% corresponding to a final temperature of 60°C. Derived constants for each wavelength are shown as well.*

Table 6: *Time constants for wild type and helix three mutant.*[34]

Conclusion

Developing a new methodology for monitoring the kinetics of ultrafast folding proteins can be used to help understand the basics of how proteins fold. By better understanding these underlying principles of folding, experimental techniques and theoretical techniques can both be improved in order for use on larger protein systems. The information that can be derived from understanding how these proteins fold will also lead to an understanding of what is different when a protein misfolds. By understanding what causes a protein to misfold, therapeutic techniques can be developed to correct for protein misfolding pathways. This has powerful implications in the treatment of diseases that are caused by misfolded proteins.

The optimized protein expression and purification protocol described above yields a proven technique to insert an infrared sensitive label into a specific site of a protein. The characterization of the protein showed that the use of minimal media to insert a heavy isotope backbone label was successful. However, there were a number of difficulties that were overcome in order to have a clean protein sample that can be used in infrared spectroscopy. After cleavage of the purification tag and removal of residual contaminates, the infrared spectrum of the proteins showed that both the buried and solvated helix bands at 1648 cm^{-1} and 1632 cm^{-1} respectively were clearly resolved. The absorbances were then used to determine a melt temperature through infrared spectroscopy. The values tracked closely with those determined by circular dichroism. Also shown in the infrared spectrum was the labeled peak at around 1560 cm^{-1}. The temperature dependence can be clearly seen showing that the label can be followed using a temperature dependent method.

The derived melt temperatures, observed cooperativity and calculated free energy values for each of the proteins showed that the helix one and helix three mutants were not significantly destabilized by introducing the methionine mutation. The helix two mutant showed a slight decrease in stability as

indicated by the lower melt temperature, but no change in the cooperativity of the folding or in the free energy of the system. This gives possible supporting evidence for the hypothesis that helix two is essential in the pathway to the native fold of BdpA. The thermodynamic analysis of the proteins shows that each of the mutant proteins is structurally similar to the wild type protein and can be used for kinetic measurements.

The preliminary kinetics showed that temperature-jump spectroscopy can be used to follow the kinetic rates of the labeled peak in the protein. The ability to monitor this signal allows for the determination of the kinetics at that particular site in the protein. The measured kinetics suggest that helix three might be present in the intermediate state, but further analysis will be required to determine the folding pathway of the protein.

The effective labeling of BdpA provides three mutant versions of the protein that can be monitored to determine the kinetics of each of the helices. The established protocol can be used to introduce a label into any position in a protein where the mutation won't cause a significant change to the protein's stability and folding pathway. The determination of the kinetics of each labeled position will give a clear picture of when that structural feature is formed in the pathway. For BdpA, the three mutants can therefore be used in kinetic experiments to determine when each of the helices form. This information will indicate which of the helices are present in the proposed .intermediate state. The successfully expressed proteins will be able to give additional experimental information for use in determining which of the proposed folding pathways for BdpA is in fact correct. Also, the developed methodology can be adapted for other proteins, providing a powerful analytical technique for determining the folding pathway of a protein.

References

[1] C. Soto, *Nat Rev Neurosci* **2003**, *4*, 49-60.

[2] K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, *Annu Rev Biophys* **2008**, *37*, 289-316.

[3] C. M. Dobson, *Nature* **2003**, *426*, 884-890.

[4] R. Vidal, B. Caballero, A. Couve, C. Hetz, *Curr Mol Med* **2011**, *11*, 1-12.

[5] D. C. Koboldt, http://www.massgenomics.org/2011/02/a-promising-new-drug-for-cystic-fibrosis.html. **2009**.

[6] R. B. Dyer, *Curr Opin Struct Biol* **2007**, *17*, 38-47.

[7] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, V. A. Voelz, *Curr Opin Struct Biol* **2007**, *17*, 342-346.

[8] H. S. Chan, Z. Zhang, S. Wallin, Z. Liu, *Annu Rev Phys Chem* **2011**, *62*, 301-326.

[9] B. Noltig, *Protein Folding Kinetics: Biophysical Methods*, Springer-Verlag, Berlin, **1999**.

[10] E. M. Boczko, C. L. Brooks, 3rd, *Science* **1995**, *269*, 393-396.

[11] V. S. Pande, *Phys Rev Lett* **2010**, *105*, 198101.

[12] C. D. Snow, E. J. Sorin, Y. M. Rhee, V. S. Pande, *Annu Rev Biophys Biomol Struct* **2005**, *34*, 43-69.

[13] K. W. Plaxco, K. T. Simons, D. Baker, *J Mol Biol* **1998**, *277*, 985-994.

[14] G. S. Buchner, R. D. Murphy, N. V. Buchete, J. Kubelka, *Biochim Biophys Acta* **2011**, *1814*, 1001-1020.

[15] H. Huang, E. Ozkirimli, C. B. Post, *J Chem Theory Comput* **2009**, *5*, 1301-1314.

[16] H. Lei, C. Wu, Z. X. Wang, Y. Zhou, Y. Duan, *J Chem Phys* **2008**, *128*, 235105.

[17] D. Mi, W. Q. Meng, Y. Q. Sun, *Phys Rev E Stat Nonlin Soft Matter Phys* **2011**, *83*, 041901.

[18] S. Sato, A. R. Fersht, *J Mol Biol* **2007**, *372*, 254-267.

[19] S. Sato, T. L. Religa, V. Daggett, A. R. Fersht, *Proc Natl Acad Sci U S A* **2004**, *101*, 6952-6956.

[20] Y. Bai, A. Karimi, H. J. Dyson, P. E. Wright, *Protein Sci* **1997**, *6*, 1449-1457.

[21] M. C. Baxa, K. F. Freed, T. R. Sosnick, *J Mol Biol* **2008**, *381*, 1362-1381.

[22] M. Zamparo, A. Pelizzola, *J Chem Phys* **2009**, *131*, 035101.

[23] J. K. Myers, T. G. Oas, *Nat Struct Biol* **2001**, *8*, 552-558.

[24] D. O. Alonso, V. Daggett, *Proc Natl Acad Sci U S A* **2000**, *97*, 133-138.

[25] A. E. Garcia, J. N. Onuchic, *Proc Natl Acad Sci U S A* **2003**, *100*, 13898-13903.

[26] S. Kmiecik, A. Kolinski, *J Am Chem Soc* **2011**, *133*, 10283-10289.

[27] A. Kolinski, J. Skolnick, *Proteins* **1994**, *18*, 353-366.

[28] A. Ghosh, R. Elber, H. A. Scheraga, *Proc Natl Acad Sci U S A* **2002**, *99*, 10394-10398.

[29] Z. Guo, C. L. Brooks, 3rd, E. M. Boczko, *Proc Natl Acad Sci U S A* **1997**, *94*, 10161-10166.

[30] S. Cheng, Y. Yang, W. Wang, H. Liu, *J Phys Chem B* **2005**, *109*, 23645-23654.

[31] A. R. Fersht, V. Daggett, *Cell* **2002**, *108*, 573-582.

[32] R. B. Dyer, D. M. Vu. Comprehensive Biophysics, *Vol 3*, The Folding of Proteins and Nucleic Acids. Oxford: Academic Press, **2012**, 34-42.

[33] K. Todar, http://www.textbookofbacteriology.net/staph_2.html. **2011**.

[34] D. M. Vu, J. K. Myers, T. G. Oas, R. B. Dyer, *Biochemistry* **2004**, *43*, 3582-3589.

[35] B. W. Caughey, A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes, W. S. Caughey, *Biochem* **1991**, *30*, 7672-7680.

[36] R. B. Dyer, F. Gai, W. H. Woodruff, *Acc Chem Res*, **1998**, *31*, 709-716.

[37] A. Barth, *Biochem et Biophys Acta – Bioenergetics*, **2007**, *1767*, 1073-1101.

[38] C. Kolano, J. Helbing, M. Kozinski, W. Sander, P. Hamm, *Nature*, **2006**, *444*, 469-472.

[39] M. T. Zanni, R. M Hochstrasser, *Curr Opinion in Struc Bio*, **2001**, *11*, 516-522.

[40] J. Kong, S. Yu, *Acta Biochim et Biophys Sinica*, **2007**, *39*, 549-559.

[41] Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, A. Tokmakoff, *Acc Chem Res*, **2008**, *41*, 432-441.

[42] N. Sreerama, R. W. Woody, *Analy Biochem*, **2000**, *287*, 252-260.

[43] J. T. Pelton, L. R. McLean, *Analy Biochem*, **2000**, *277*, 167-176.

[44] D. Seebach, R. E. Ciceri, M. Overhand, B. Juan, D. Rigo, L. Oberer, U. Hommel, R. Amstutz, H. Widmer, *Helv Chim Acta*, **1996**, *79*, 2043-2066.

[45] W. C. Johnson, *Proteins: Struc, Func, and Bioinform*, **1999**, *35*, 307-312.

[46] J. Marecek, B. Song, S. Brewer, J. Belyea, R. B. Dyer, D. P. Raleigh, *Org Lett*, **2007**, *9*, 4935-4937.

[47] S. H. Brewer, B. Song, D. P. Raleigh, R. B. Dyer, *Biochemistry* **2007**, *46*, 3279-3285.

[48] J. Marecek, B. Song, S. Brewer, J. Belyea, R. B. Dyer, D. P. Raleigh, *Org Lett* **2007**, *9*, 4935-4937.

[49] H. Taskent-Sezgin, J. Chung, P. S. Banerjee, S. Nagarajan, R. B. Dyer, I. Carrico, D. P. Raleigh, *Angew Chem Int Ed Engl* **2010**, *49*, 7473-7475.

[50] R. Callender, R. B. Dyer, *Curr Opin Struct Biol* **2002**, *12*, 628-633.

[51] J. D. Young, A. S. Huang, N. Ariel, J. B. Bruins, D. Nq, R. L. Stevens, *Pept Res*, **1990**, *3*, 194-200.

[52] K. Noren, J. S. Loring, J. R. Bargar, P. Persson, *J Phys Chem*, **2009**, *113*, 7762-7771.

[53] N. Greenfield, *Nat Protoc*, **2006**, *1*, 2527-2535.

[54] S. Nagarajan, H. Taskent-Sezgin, D. Parul, I. Carrico, D. P. Raleigh, R. B. Dyer, *J Am Chem Soc*, **2011**, *133*, 20335-20340.

[55] H. Susi, S. N. Timasheff, L. Stevens, *J Biol Chem* **1967**, *242*, 5460-5466.

[56] L. Tadesse, R. Nazarbaghi, L. Walters, *JACS*, **1991**, *113,* 7037-7039.

[57] J. Hofrichter, *Methods Mol Biol* **2001**, *168*, 159-191.

[58] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, R. B. Dyer, *Biochemistry* **1996**, *35*, 691-697.

[59] F. Huang, E. Lerner, S. Sato, D. Amir, E. Haas, A. R. Fersht, *Biochemistry* **2009**, *48*, 3468-3476.

[60] J. Werner, R. B. Dyer, R. M. Fesinmeyer, N. H. Andersen, *J Phys Chem B* **2002**, *106*, 487-494.