The Role of a Conserved Helix in the Structural Evolution of Cro Protein

Tiffany G. Son

A Thesis Submitted to the Honors College
In Partial Fulfillment of the Bachelors Degree with Honors in
Biochemistry and Molecular Biophysics

Department of Biochemistry & Molecular Biophysics
THE UNIVERSITY OF ARIZONA
May 2011

Approved By:

Dr. Matthew H.J. Cordes, Principle Investigator — Date: 04.15.2011

Dr. Tsu-Shuen Tsao, Undergraduate Faculty Advisor — Date: 4/22/2011

Table of Contents

Statement by Author

I hereby grant to the University of Arizona Library the nonexclusive worldwide right to reproduce and distribute my thesis and abstract (herein, the "licensed materials"), in whole or in part, in any and all media of distribution and in any format in existence now or developed in the future. I represent and warrant to the University of Arizona that the licensed materials are my original work, that I am the sole owner of all rights in and to the licensed materials, and that none of the licensed materials infringe or violate the rights of others. I further represent that I have obtained all necessary rights to permit the University of Arizona Library to reproduce and distribute any nonpublic third party software necessary to access, display, run, or print my thesis. I acknowledge that University of Arizona Library may elect not to distribute my thesis in digital format if, in its reasonable judgment, it believes all such rights have not been secured.

SIGNED: 

Abstract

Proteins are diverse in both structure and function, allowing for the inquiry of how these differences evolved. The Cro protein family contains two prophage proteins, Xfaso 1 and Pfl 6, that have evolved different folds. Xfaso 1 has an all α helical structure and Pfl 6 has an α+β fold. One goal of our research is to determine which features of the Xfaso 1 and Pfl 6 sequence are critical for the fold each protein adopts, to illuminate possible mechanisms leading to structural change. This study focuses on residues located on the interior of helix 1, which is conserved in both structures, but interacts closely with the divergent C-terminal region of the protein. Due to variable size and shape of these residues found between Xfaso 1 and Pfl 6, the packing interface with the C-terminal region is quite different. We made three single mutations in helix 1 of Pfl 6 involving introduction of single residues from Xfaso 1: Ile6Leu, Tyr9Ala, His13Leu; we also constructed the triple mutant Leu6Ile/Tyr9Ala/His13Leu. Additionally, three single mutations were made in Xfaso 1 involving the single residues from Pfl 6: Leu4Ile, Ala7Tyr, and Leu11His. Each residue's importance was examined by measuring effects on structure and stability to investigate if the mutations lead to conversion from an α+β fold to an all α-helical fold. Circular dichroism monitored two measurements: the secondary structure with the scans in the far ultraviolet range and the stability of the mutated proteins through the use of thermal melts. Preliminary results show no evidence for conversion to a helical fold but show some effects on stability of the α+β fold. Tyr9Ala is strongly destabilizing while His13Leu is slightly stabilizing and Leu6Ile is slightly destabilizing. In this interface, the determining factors for stability in the Cro folds are the size and hydrophobic nature of the residues.

Introduction

Proteins play a pivotal role in all organisms which vary from the cell membrane where they are channel or carrier proteins, hormones, immuno-proteins, transporters, and enzymes. An incredible diversity of protein domains, more than 1,000 different topologies, is constructed from the same 20 amino acids and the same few repeating structural elements to give every protein a three-dimensional fold. Some folds have a superficial similarity, like the immunoglobulin fold and the gamma crystallin like fold, but are very different upon close inspection. There is some, but not a perfect correlation of fold with function. TIM barrels are often enzymes, SH2 domains have a signaling function, globulins frequently bind heme, gamma crystallins form aggregate structures like those found in the human eye lens. A small non-representative subset is shown in Figure 1 (Figure Courtesy of Matthew Cordes).

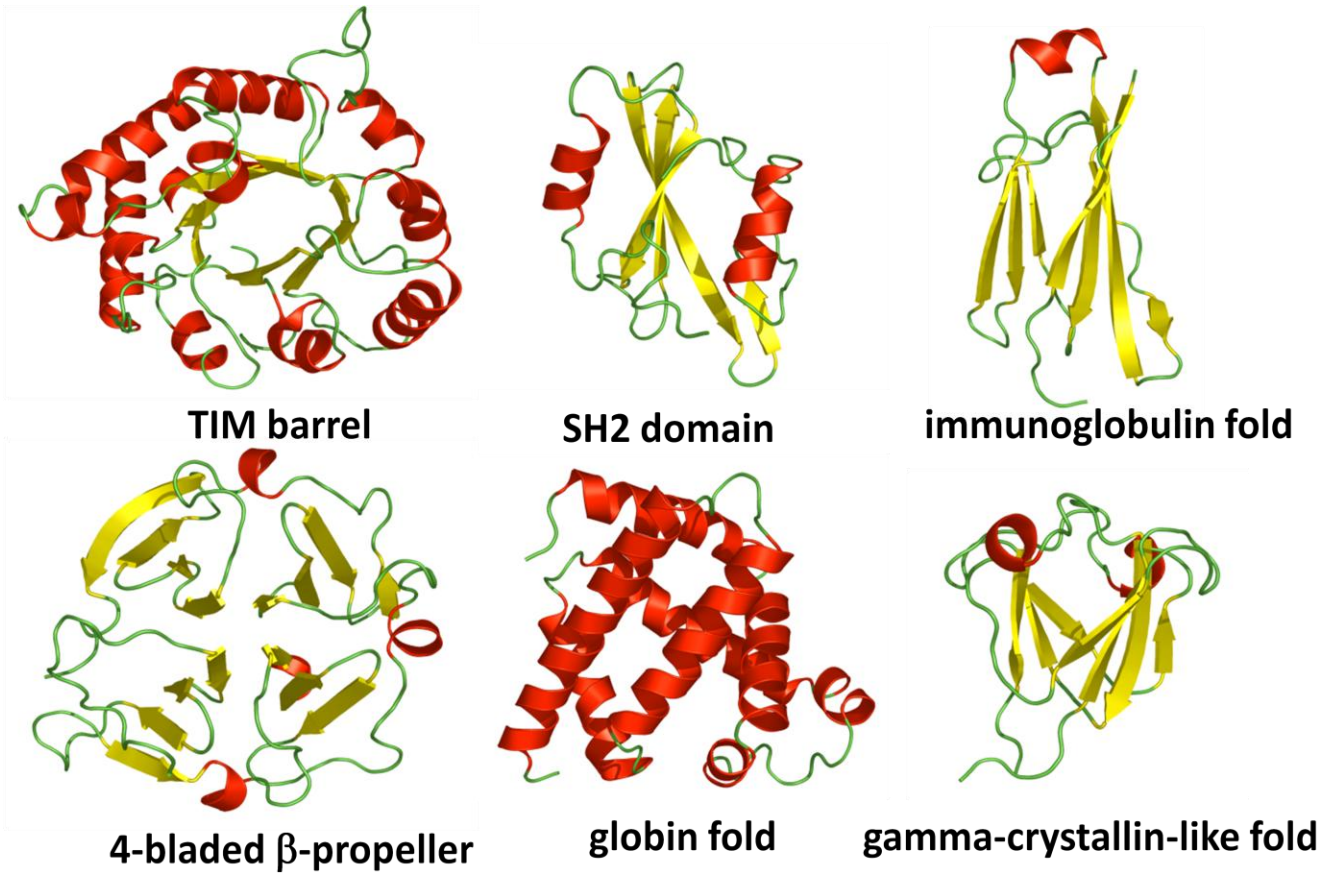

Figure 1: Diversity of Protein folds. A limited depiction of structural topologies of proteins that encompass a broad diversity (Courtesy of Matthew H.J. Cordes).

These basic classifications of protein domains raise the question of whether or not some protein structures evolved by modification of pre-existing structures through mutations in the primary amino acid sequence (Cordes et al. 1999; Grishin 2001).

Although the amino acid sequence has influence on the fold of the protein since it encodes all the information for a functional structure (Anfinsen, 1973), the sequence governs the fold in a global, complex manner, and very similar sequences can adopt different folds. Previous experiments have shown that the outcome of the protein fold may be influenced by the location of the key amino acids in the sequence and the interactions that take place to create α-helical structure or β-sheet structures (Minor and Kim, 1996). The mutational mechanism in which the evolutionary adaptation occurs is studied in proteins with three different criteria: common ancestor, similar amino acid sequence, and different structural folds. The Cro protein family fits all three criteria.

Divergent evolution has played a pivotal role in the diversity of the Cro protein superfamily. Through changes in amino acid sequences, this type of evolution has led to structural as well as functional differences that question the evolutionary process of preexisting folds to various secondary structures. The Cro protein specified by bacteriophage lambda is a repressor of the genes expressed early in phage development and is required for a normal late stage of lytic growth (Biology Hypertextbook Index). The lytic pathway differs in the bacterial cell compared to the lysogenic pathway by going straight to lysis of the cell and release of new phages. On the other hand, the lysogenic pathway comes to latency by the phage DNA integrating with the bacterial chromosome. When the bacteria have no environmental stress factors and have many bacteria to infect

around then, the phage will choose to replicate during this advantageous stage. However, if there are environmental stress factors and little potential for bacterial growth, the phage will preferentially go into the lysogenic state. The decision between these two fates is decided by two competing proteins in the lambda bacteriophage, CI and and Cro (Biology Hypertextbook Index). CI promotes the lysogenic cycle while the Cro protein promotes the lytic phase. These two proteins are in direct competition with each other for sites on the promoter region of lambda. When Cro is transcribed and translated, the lytic pathway is activated and Cro transcriptionally represses CI (Figure 2B). When CI is transcribed and translated, the lysogenic pathway is activated and CI transcriptionally represses Cro (Figure 2A).

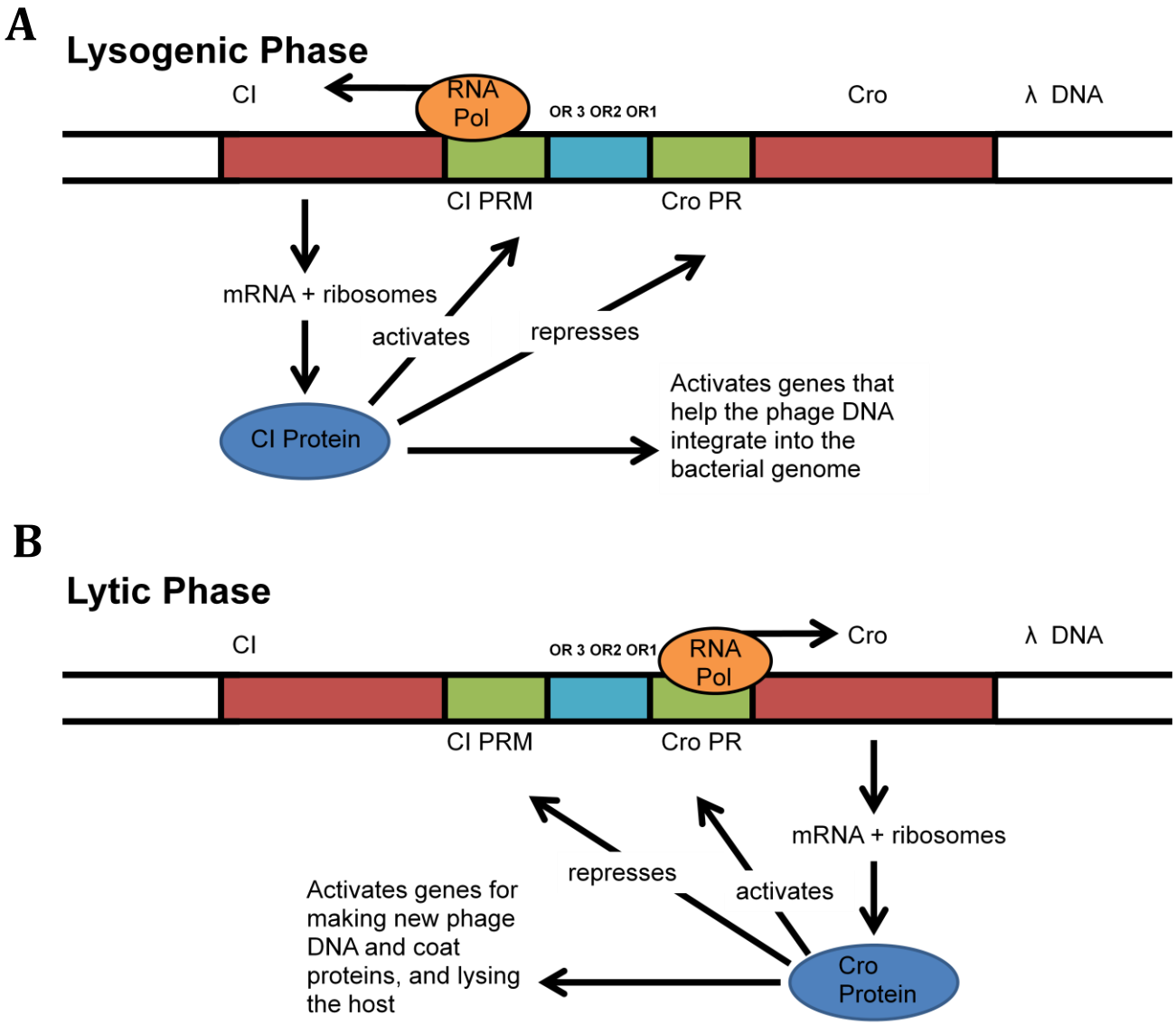

Figure 2: The Lambda Phage Cycle Decision. CI starts to win the competition when the bacteria are not growing very well, because it is continuously being made at a low level. When the bacteria are not dividing as often, the concentration builds up and CI inhibits the lytic phase (Adapted from Biology Hypertextbook Index).

The Cro and CI proteins have an all α-helical structure composed of five to six α-helices but some members of the Cro Family have evolved a change from all α-helical structure to a mixed α+β fold. In the Cro protein family, P22 Cro represents the ancestral all helical structure of the Cro protein family and λ Cro represents the descendant mixed α/β structure (Newlove et al, 2004). Additionally, the folds of P22 Cro and λ Cro have three α-helices that make up the N-terminal half of the domain, but differ in structure in the C-terminal half of the domain. For these two proteins a transitive homology model (Figure 3) aids in the explanation of this transition through three intermediate Cro sequences; Cro proteins that are evolutionarily between P22 Cro and λ Cro (Roessler et al., 2008).

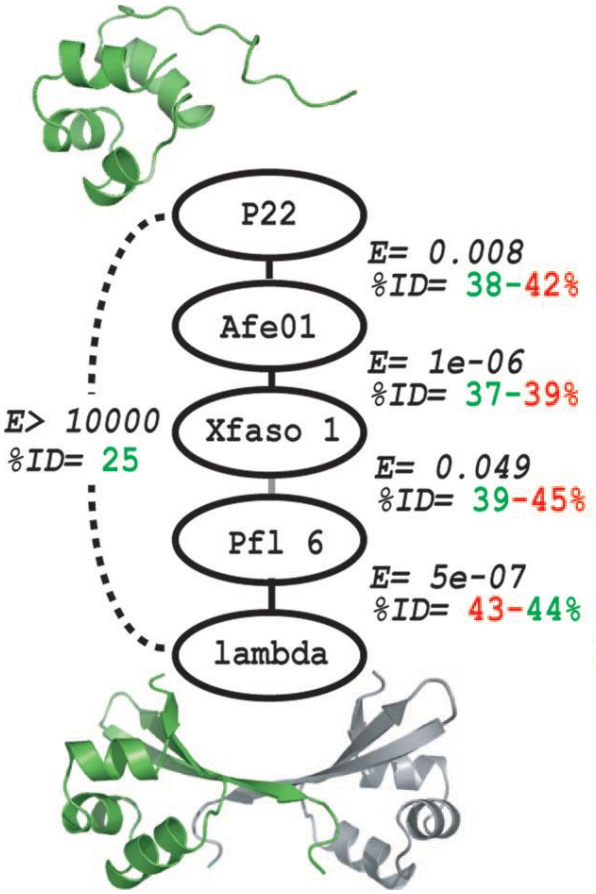

Figure 3: Transitive Homology Analysis. Three prophage Cro proteins, Afe01, Xfaso 1, and Pfl 6 were used as sequence intermediates in a transitive homology analysis connecting P22 Cro to λ Cro. P22 Cro is compromised of an all α-helical fold, while λ Cro is comprised of a mixed α+β fold. Xfaso 1, Pfl 6, and Afe01 are intermediate in sequence to P22 Cro and λ Cro. Percent sequence identities (%ID) are given for each link in the chain (adapted from Roessler et al. 2008).

Through the use of transitive homology analysis, two particular protein intermediates in the sequence between P22 Cro and λ Cro were selected to gain a more complete understanding of the mechanism by which a protein may evolve into a new type of structure via mutations in its amino-acid sequence. In the Cro protein family, two prophage species with different topologies are examined closely, *Pseudeomonas fluorescens* (Pfl 6) and *Xylella fastidiosa* (Xfaso 1) (Roessler, et al. 2008). Pfl 6 and Xfaso 1 mirror the evolutionary structural switching of P22 Cro and λ Cro since Xfaso 1 has an all α-helical fold while Pfl 6 has an α+β fold. Between these two proteins there is a 40% sequence identity whereas P22 Cro and λ Cro have a 22% sequence identity. Through the use of x-ray crystallography, the atomic resolution structures of these proteins were solved (Figure 4) (Roessler, et al. 2008). While the N terminus of both these proteins is structurally conserved, the C terminus is structurally diverged.

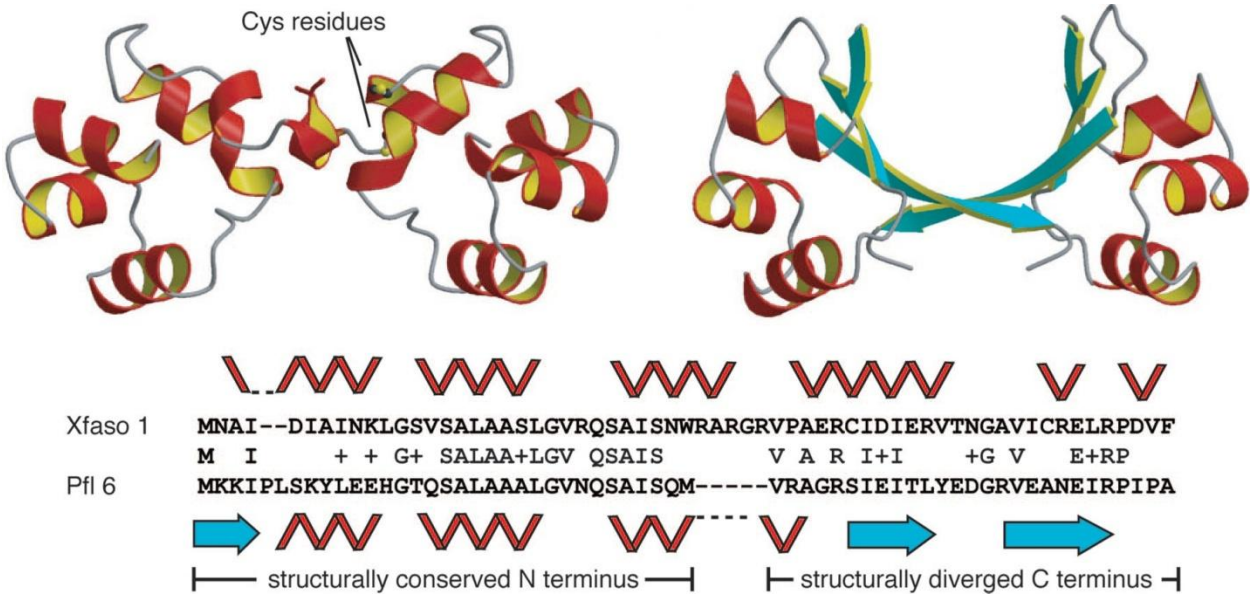

Figure 4: Pfl 6 and Xfaso 1. Homologous proteins with 40% sequence identity and different folds (Roessler et al. 2008).

Although sequence differences in the C-terminal half are important, chimera mutagenesis and design studies show that they are insufficient by themselves to account for the differences in structure (Karen V. Eaton, William J. Anderson, Emily M. Dykstra, Matthew H.J. Cordes, unpublished data). The C-terminal structure is determined not just by its own sequence but by how this sequence interacts with the N-terminal half since in the three dimensional structure the difference is in the central fold. It is therefore of interest to study the effect of differences in the N-terminal halves of the two sequences.

This particular study focuses on residues located on the interior of helix 1, which is conserved in both structures, but interacts closely with the divergent C-terminal region of the protein. One possibly important factor in selecting one C-terminal structure over the other is how helix 1 in the N-terminal region packs against the C-terminal structures.

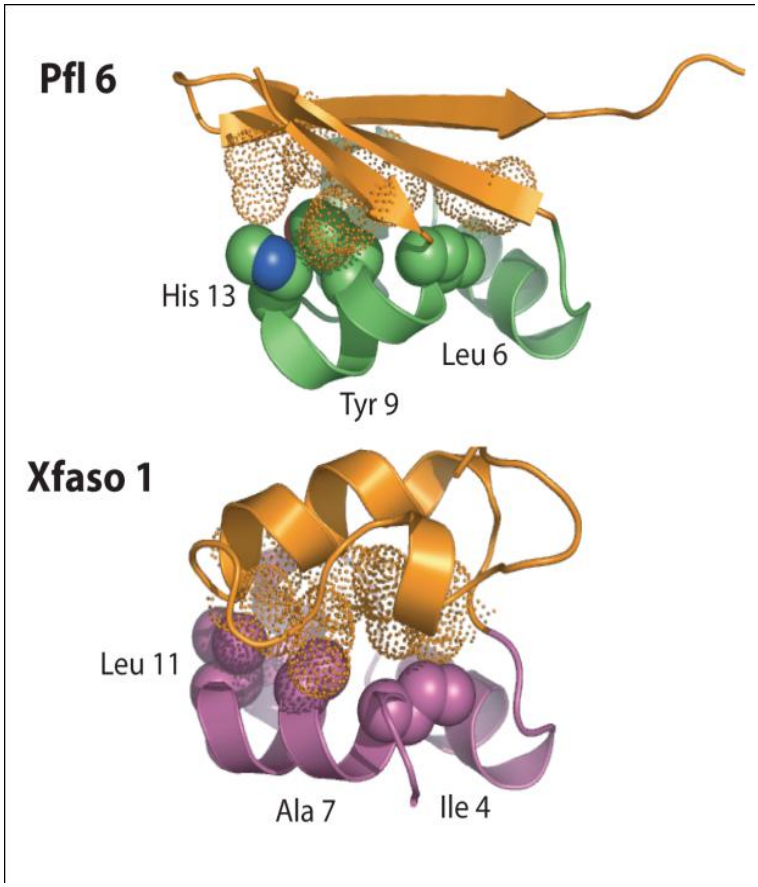

Figure 5: Interactions between the N-terminus and the C-terminal structures of Pfl 6 and Xfaso 1. The C-terminus is in gold and the N-terminus in purple or green. The three major packing side chains from Pfl 6 are Leu 6, Tyr 9 and His 13 and for Xfaso 1 they are Ile 4, Ala 7, Leu 11. Interacting side chains from the C-terminus are shown in dotted spheres (Courtesy of Matthew H.J. Cordes).

Figure 5 shows that helix 1 has a different angle in Pfl 6 and Xfaso 1 with respect to the rest of the protein, and it also has different side chains that pack against the C-terminal amino acid side chains. The three major packing side chains from Pfl 6 (Leu 6, Tyr 9 and His 13) have higher volume than those from Xfaso 1 (Ile 4, Ala 7, Leu 11), largely due to the difference between Tyr and Ala at the central position. The bowed-out appearance of helix 1 in Pfl 6 may be due to this higher packing volume requiring more space.

We designed three single mutations in helix 1 of Pfl 6 involving introduction of single residues from Xfaso 1; we also constructed the triple mutant. Also, we made three single mutations in helix of Xfaso 1 involving introduction of single residues from Pfl 6. Each residue's importance was examined by measuring effects on structure and stability to investigate if the mutations lead to conversion from an α+β fold to an all α-helical fold. Accordingly, this project plans to examine each individual amino acid change in that region in order to determine how each of these amino acids plays a role in the overall stability and structure of the protein.

The evolutionary process seen in Xfaso 1 and Pfl 6 is drastic, which may imply that evolution has favored an all or nothing process. One of the possible expected outcomes of this project is the conclusion that all the amino acids in this region are important and exchanges of residues between the two proteins will destabilize the protein. Also, the possibility exists that only one or two amino acids in this region are critical while the other amino acids cause no effect. The change of one amino acid to completely modify the structure of the protein is seen as highly improbable since multiple changes as well as

interactions are needed for this extreme change to occur. We may hypothesize that substitution of the Xfaso 1 side chains for those of Pfl 6, or vice versa, one at a time or all together, will either destabilize one or both proteins or alter the conformation of helix 1. An improved understanding of the origin and the evolution of protein folds would help develop the evolutionary steps taken to create a diverse population of proteins with various topologies.

Methods and Materials

QuikChange site-directed mutagenesis

The mutagenic oligonucleotide primers used were designed individually to contain the desired mutation or mutations and to anneal to the same sequence on opposite strands of the plasmid template. Certain mutations were introduced using multiple sequential mutagenesis primers. The primers were designed to have a T_m of ≥78°C using the following formula as an estimation of the T_m for the primers:

$$(1) \quad T_m = 81.5 + 0.41(\%GC) - \frac{675}{N} - \%\, mismatch.$$

N is the primer length in bases and the values for %GC and % mismatch are whole numbers.

These desalted custom synthesized DNA oligos were purchased from Integrated DNA Technologies (IDT). Mutations were introduced into pET21b-based (Novagen) plasmids containing genes for either wild type Pfl 6 Cro (Pfl 6 I58D, a monomeric variant of wild type Pfl 6) or Xfaso 1 Cro (pWA100) and the original constructs of Pfl 6 and Xfaso 1 had a 6 histidine tag to aid in purification and to prevent proteolytic degradation, which were retained in the mutants. The two synthetic oligonucleotide primers contained the hybrid mutations in Pfl 6 individually (Leu6Ile; Tyr9Ala; and His13Leu) and together (Leu6Ile/Tyr9Ala/His13Leu) and Xfaso 1 hybrid mutations individually (Ala7Tyr; Ile4Leu; and Leu11His).

The mutant strand synthesis reaction contained: 5 µL of 10x Pfu Ultra AD buffer; 2.5 µL of each complimentary primer (5 µM); 1 µL dNTP mix; 2 µL of each respective plasmid template (10 µg/µL); and 37 µL of DI water. Then 1 µL of Pfu Ultra DNA polymerase (2.5

U/µL) was added to the reaction and underwent 18 cycles using an *Eppendorf Mastercycler PCR thermal cycler*. The thermal cycling parameters included an initial denaturation step that involved heating the reaction to 95˚C for 30 seconds (denaturation), cooling to 55˚C for 1 minute (annealing), then heating to 68˚C for 1 minute per kilobase of template plasmid length (extension). Following the temperature cycling, 1 µL of *Dpn* I restriction enzyme (10 U/µL) was added directly to each amplification reaction and incubated at 37°C for 1 hour to digest the parental DNA. This amplification product was then analyzed by electrophoresis of 10 µL of the product on a 1% agarose gel.

Verification through Sequencing

Then, 1 µL of the solution of mutated plasmid created through mutagenesis thermal cycling was then transformed into 100 µL XL1-Blue $CaCl_2$ competent *E. coli* cells through heat shock transformation. After the outgrowth of the transformation mixtures in 1 mL of Luria-Bertani (LB) broth for 45 m, 100 µL of the mixture was spread on an LB agar plate with 100 µg/mL ampicillin (Amp). Plates were incubated overnight at 37˚C and then colonies were picked the after 12-16 hours and grown overnight in 5 mL LB containing 100 µg/mL Amp. The selected plasmid DNA was purified through the *QIAprep Spin Miniprep Kit* (Qiagen). Concentrations of purified plasmids were determined from A_{260} values using a NanoDrop 2000c (Thermo Scientific). The following equation was used to determine concentrations from the A_{260} values:

(2) $$C(\mu g/mL) = \frac{OD_{260}}{0.02}$$

The presence of the desired mutation or mutations was verified by capillary electrophoresis sequencing through the University of Arizona Genetics Core facility with 5 μL of a 200 μg/mL plasmid sample.

Protein Expression and Purification

Upon correct sequence verification, the modified plasmid containing the desired mutation was transformed into BL21 (λDE3) $CaCl_2$ competent *E. coli* cells according to standard $CaCl_2$ transformation protocol. A single colony was picked and the culture was grown in 40 mL LB-amp with 250 rpm shaking at 37°C until OD_{600}= 0.55 at which point protein expression was then induced with 100 μg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown in the same conditions for 3 additional h. All spectrophotometric analysis was completed on a Cary 50 UV-Vis spectrophotometer. The culture was then harvested by centrifugation at 5,000 × g at 4˚C for 5 m. The pellets were resuspended in 0.6 mL Ni-NTA Buffer A (100 mM NaH_2PO_4[pH 8.0], 10 mM Tris, 6M guanidine-HCl) in order to eliminate inclusion bodies. Buffer A also included a $\frac{1}{1000^{th}}$ volume of β-mercaptoethanol (BME) for disulfide reduction in Xfaso 1 constructs. The suspension was inverted for 1 hour at 4˚C to complete lysis and then centrifuged for 15 minutes at top speed in a microcentrifuge. The supernatant was then brought to 10 mM imidazole concentration by adding the appropriate volume of buffer A + 250 mM imidazole.

The lysate was loaded directly to a Ni-NTA affinity spin column (Qiagen) which had been equilibrated with 0.6 mL buffer A + 10 mM imidazole. After the lysate had been loaded, the spin columns were washed twice with 0.6 mL buffer A+ 10 mM imidazole, each time spinning 2 minutes at 790 × g. The wash steps were done to remove nonspecifically bound

impurities. The protein was eluted from the column with two 0.3 mL fractions of Buffer F (0.2 M acetic acid, 6 M guanidine-HCL) that were later combined.

The eluate was dialyzed in 1 × SB250 (50 mM Tris [pH 7.5], 250 mM KCl, 0.5 mM EDTA) for 48 h while changing to a fresh dialysis solution after the first 24 hours. For Xfaso 1 proteins, the eluate was dialyzed with 1mM dithiothreitol (DTT) in 1 × SB250.

The purity of the dialyzed protein was analyzed by electrophoresis on Tris-Tricine SDS-PAGE gels. The sample buffer (3.15 mL 1M Tris [pH 6.8], 29.4 mL H_20, 10 mL 10% sodium dodecyl sulfate (SDS), 5 mL glycerol, 2.5 mL BME, and a dash of bromophenol blue) of 20 µL was added to 20 µL protein sample and heated in a H_20 bath for 10 minutes at 95°C. The ladder (5 µL sample buffer and 5 µL SDS-Page broad range standards) was also heated by the same method. The gel consisted of a separating gel and a stacking gel. A gel buffer (3 M Tris (pH 8.45), 0.3% SDS) was utilized in both parts. The separating gel was made with 5 mL ProtoGel (30% acrylamide, 0.8% bis-acrylamide; National Diagnostics), 5 mL separating buffer (200 mL gel buffer, 63.5 mL glycerol, 0.2 µL filtered), 33µL 10% ammonium persulfate (APS), and 3.3 µL tetramethylethylenediamine (TEMED). The stacking gel was made with 0.64 mL ProtoGel, gel buffer, 3.1 mL deionized H_20, 40 µL 10% APS, and 4 µL TEMED. After the gel ran at 120 V for approximately 1.5 h, the gel was stained for an hour with Coomassie R-250 Brilliant Blue stain (464 mL ethanol, 45 mL acetic acid, 464 mL H_20, 2.4 g Coomassie blue R-250) and later put into destaining solution (300 mL ethanol, 92 mL acetic acid, 608 mL H_20).

Spectrophotometric Calculation of Protein Concentration

After dialysis, the protein samples were centrifuged at 14,000 rpm in a microcentrifuge for 20 m to remove any precipitates. A working concentration of protein for circular dichroism was prepared by diluting the protein to optimally give an A_{280} near 0.5. Then a spectrum from 220 to 340 nm was taken of the solution against the correct blank, SB250 dialysis buffer. The following formula was used to obtain the corrected A_{280} by taking into account the wavelength dependent light scattering effects due to particulates.

$$(3) \quad A_{280\,corrected} = A_{280} - [(|A_{320} - A_{340}| \times 3) + A_{340}]$$

Protein concentrations were determined based on tryptophan and tyrosine absorbance at 280 nm. The extinction coefficient at 280 nm (ε_{280}) for each mutant protein of Xfaso 1 was calculated by adding 5,559 L $mol^{-1}cm^{-1}$ per tryptophan and 1,197 L $mol^{-1}cm^{-1}$ per tyrosine, which are estimated values based on an amino acid absorbance per tyrosine, which are estimated based on amino acid absorbance in 100 mM phosphate pH 7.1. for an extinction coefficient of 6,756 L $mol^{-1}cm^{-1}$ for the Ala7Tyr mutation and 5,559 L $mol^{-1}cm^{-1}$ for both Ile4Leu and Leu11His mutations (Edelhoch 1967; Gill and von Hippel 1989). For Pfl 6, which contains two tyrosines and no tryptophans, the extinction coefficient was experimentally favored to be 3,119 L $mol^{-1}cm^{-1}$ per two tyrosines in both Leu6Ile and His13Leu (Roessler et al. 2008) using the Edelhoch method (Edelhoch 1967; Gill and von Hippel 1989). The extinction coefficient for Pfl 6 Tyr9Ala has one tyrosine and was experimentally favored to be 1,559.5 L $mol^{-1}cm^{-1}$. The protein concentration was then calculated using Beer's law:

$$(4) \quad C = \frac{A_{280\,corrected}}{\varepsilon} \times dilution$$

Measured purified protein concentrations from various preparations spanned a range of ~70-100 μM.

Circular Dichroism

25 μM protein samples of 600 μL volume were subjected to circular dichroism wavelength scans and thermal denaturation monitored at 222 nm. These circular dichroism experiments were completed on the Olis® *DSM · 20 CD spectrophotometer*. The far-UV samples used a 1.0 mm pathlength cuvette and the thermal melt samples used a 2.0 mm path length cuvette. The far UV measurements required a baseline scan of the appropriate solution of SB250 with or without DTT depending on the presence of cysteines. The far UV scans were monitored from 205 nm to 260 nm at 20˚C and an averaging of 3 scans with 10 s integration time. Also, the thermal melts measured the proteins' ellipticity at a constant wavelength of 222 nm over a temperature range from 20˚C to 80˚C with a 2 minutes temperature equilibration at each temperature. An incubation time of 2 minutes was used to ensure samples were equilibrated for each ellipticity measurement.

Data Fitting of CD Thermal Melts

The graphing program Kaleidagraph (Synergy Software) was used to import and graph the CD data. Spectral and melt data were baseline corrected using spectra of SB250 buffer alone (or SB250 +1 mM DTT) and normalized base on direct spectrophotometric measurements of sample concentrations, which varied slightly around 25 μM. Melting temperatures were determined for folded variants by fitting the thermal melt data to a two-state model for the unfolding equilibrium, described by the following equation:

$$(5) \quad \Delta G_U = \Delta H_U \left(1 - \frac{T}{T_m}\right) + \Delta C_P \left(T - T_m - T ln\left(\frac{T}{T_m}\right)\right)$$

T_{m} is the melting temperature at which ΔG_U is zero, ΔH_U is the enthalpy change of unfolding at T_{m}, and ΔC_P is the heat capacity of unfolding which is assumed to be temperature invariant. The ΔC_P term was fixed at 840 cal $mol^{-1}K^{-1}$ during fitting, based on an estimate of 14 cal $mol^{-1}K^{-1}$ per residue of folded protein (Myers et al. 1995), while all other parameters were left to vary. The fitted parameters also included slopes and y-intercepts for folded and unfolded baseline, which were assumed to be linear.

Results

The main goal of this work was to examine the residues located on the interior of helix 1, which is conserved in both structures, but interacts closely with the divergent C-terminal region of the protein. The role of these residues on the overall structural characteristics as well as stability was investigated to see if the mutations are tolerated by the Xfaso 1 and Pfl 6 structures, and if any lead to conversion from an α+β fold to an all α-helical fold or vice versa.

Gel Electrophoresis Examination of Site-Directed Mutagenesis Product

After the mutagenesis reaction, successful template replication was assessed by performing agarose gel electrophoresis (1.0% in TBE) with 10 µL of the mutagenesis reaction product. An example of this confirmation for the Pfl 6 mutant Tyr9Ala and the Xfaso 1 mutant Ala7Tyr shows the correct band size of 6,200 kb.

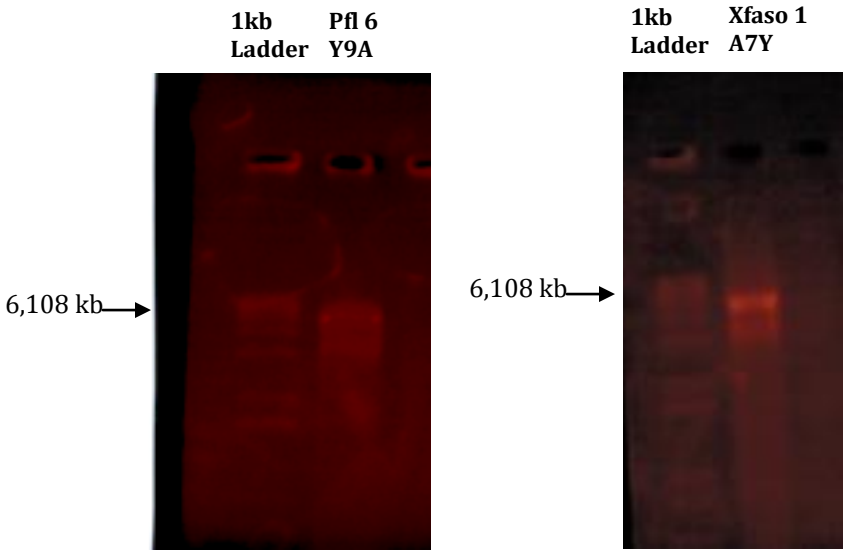

Figure 6: Monitoring with gel electrophoresis of (A) Pfl 6 mutagenesis of Tyr9Ala mutation and (B) Xfaso 1 mutagenesis of Ala7Tyr. The 1 kb ladder marker of 6,108 kb is noted.

SDS-PAGE Examination of Purity of Dialysate Protein Product

SDS-PAGE is commonly used to determine the molecular weight, concentration and relative purity of a target protein. The purified protein dialysates were also run on SDS-PAGE Tris-Tricine gels to verify purity (Figure 7).

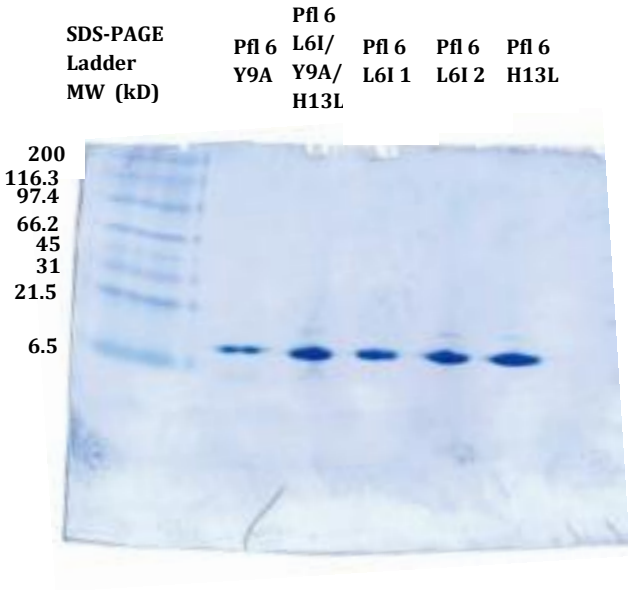

Figure 7: Coomassie-stained gel (15% SDS-PAGE) of the four purified Pfl 6 variant eluates by Ni-affinity chromatography from dialysis were seen to be pure via SDS-PAGE Tris-Tricine gel electrophoresis.

Far-UV Spectra Depicts the Mutations Effect on Overall Protein Structure.

One method to examine secondary structure is by circular dichroism in the far-UV region (190-250 nm). Far-UV wavelength scans measured by circular dichroism spectroscopy were conducted on the three single mutations of Pfl 6 (Figure 8) and Xfaso 1 (Figure 9). Also, a far-UV wavelength scan was done on a triple mutant of Pfl 6 (Figure 8). The scans measure the signal from the peptide bonds and α-helices, β-sheets, and random coils all have characteristic shapes and magnitudes in a CD spectrum. The α-helical character is particularly apparent in the minima at 208 nm and 222 nm. For an alpha helical protein with increasing amounts of random coil present, the 222 nm minimum will have a less negative ellipticity and the 208 nm minimum moves to a more negative ellipticity. Consequently a higher ratio of 208/222 for a variant relative to wild type indicates a decrease in α-helical structure and the mutations effect on the folded or unfolded state of the protein. From these 208/222 ratios, the mutations are quantitatively compared with those of the respective wild types. A low ratio indicates more folded and/or generally more helical character and a high ratio indicates less folded and/or less helical character. The CD signal reflects an average of the entire molecular population and therefore the technique is limited and cannot determine which specific residues are involved in the helical portion.

Mutations in Pfl 6

For the mutations in Pfl 6, the data depicts Leu6Ile to retain most of the α-helical structure seen by the slight shifts in ellipticity at 222 and 208nm and has a 208/222 ratio of 1.14 compared to Pfl 6-wt of 1.24. The loss of α-helical characteristics was seen in Tyr9Ala with

a 208/222 ratio of 1.6. In His13Leu mutation in Pfl 6, there is an elevation in the 208/222 ratio, 1.26, which indicates that it maintains an α-helical structure. Lastly, the triple mutations Leu6Ile/Tyr9Ala/His13Leu in Pfl 6 had a higher ratio of 1.54 compared to the wild type which depicts the loss of α-helical structure.

Mutations in Xfaso 1

Ile4Leu had a 208/222 ratio of 1.16, while the wild type of Xfaso 1 data is 1.24. The Ala7Tyr mutation in Xfaso 1 had a very high 208/222 ratio of 2.00. Another one of the Xfaso 1 reciprocal mutants was Leu11His and it had a 208/222 ratio of 1.79 which demonstrates the large change in ellipticity of this mutant protein compared to the wild type Xfaso 1 ellipticity ratio of 1.24.

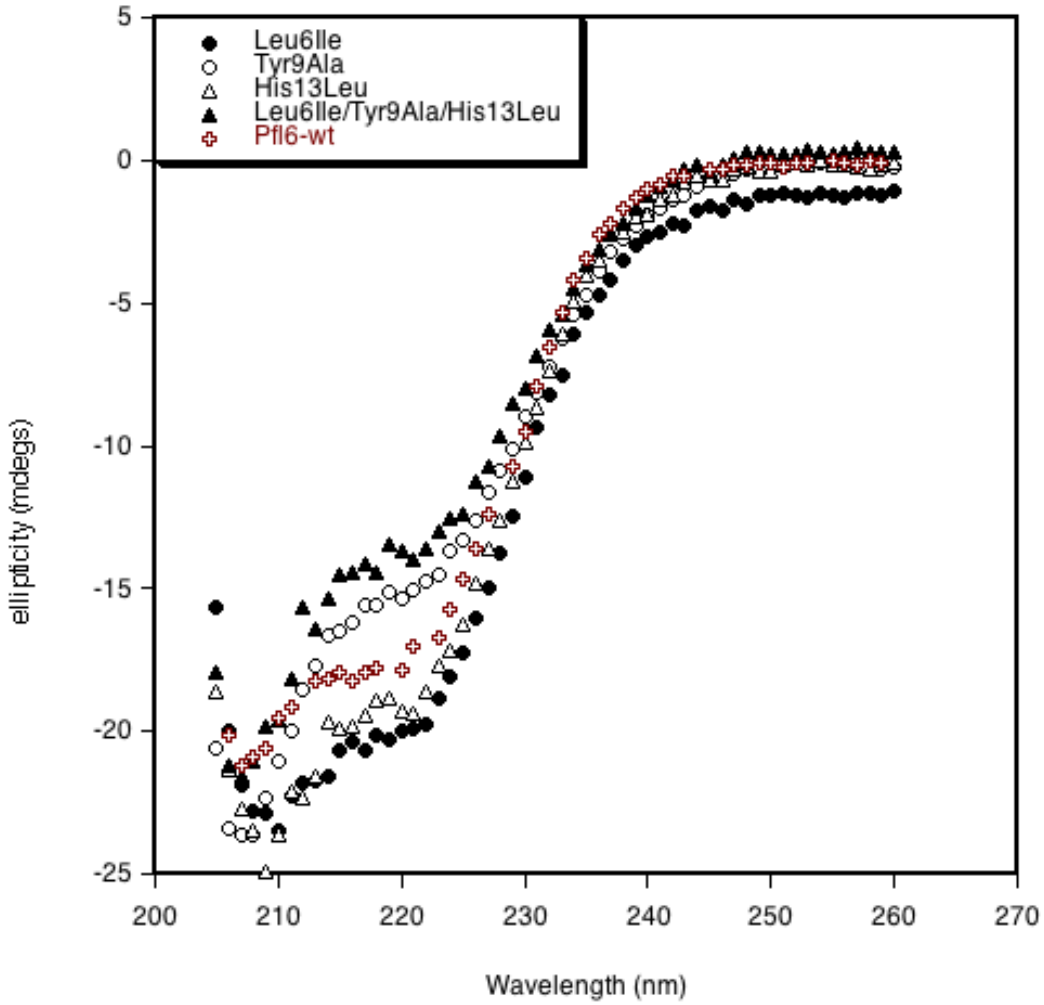

Figure 8: Far UV Spectra of Pfl 6 Mutants: 25 μM mutant Leu6Ile, Tyr9Ala, and His13Leu and the triple mutant were scanned at 20˚C from 205-260 nm in a 1 mm path length cell. All Pfl 6 protein samples were dialyzed into SB250 and spectra were both baseline-corrected with SB250.

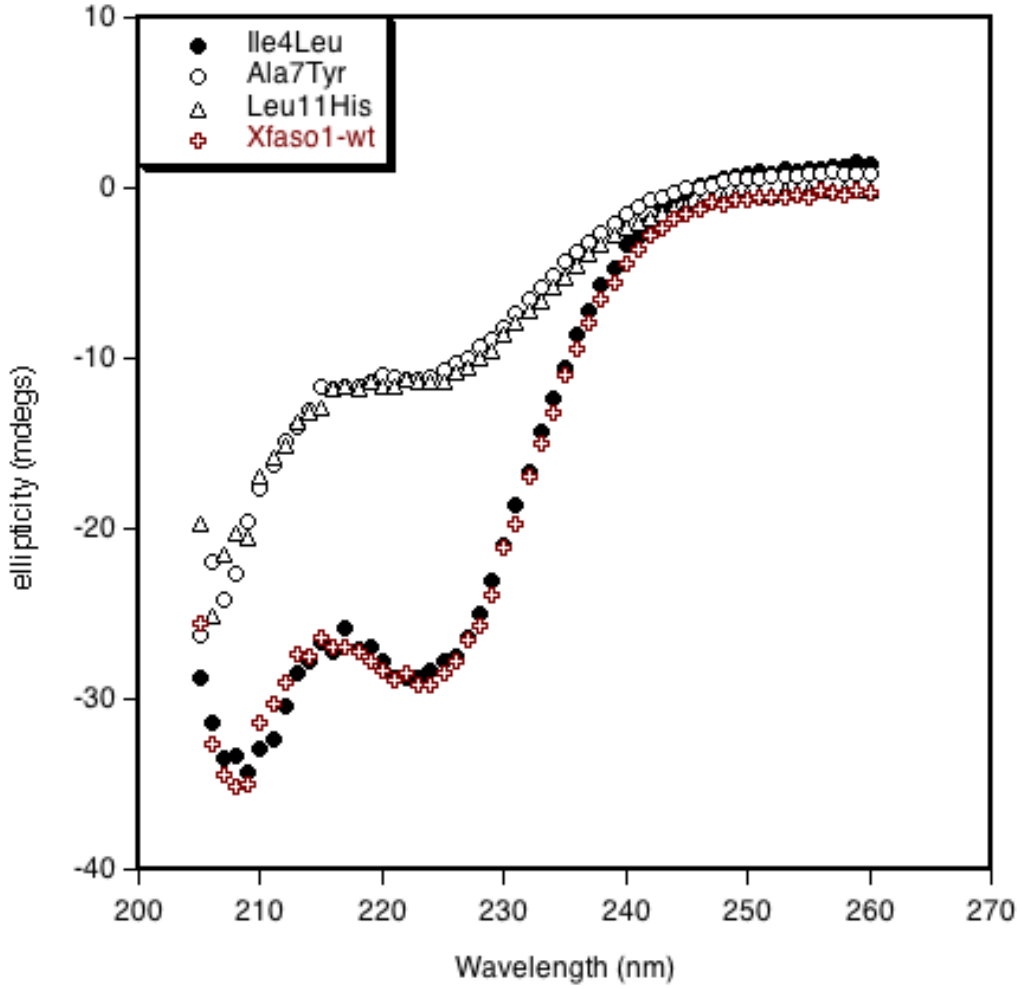

Figure 9: Far UV Spectra of Xfaso 1 Mutants: 25 μM mutant of each mutant Ile4Leu, Ala7Tyr, and Leu11His were scanned at 20˚C from 205-260 nm in a 1 mm path length cell. All Xfaso 1 protein samples were dialyzed into SB250 + 1 mM DTT and spectra were both baseline-corrected with SB250 + 1 mM DTT.

Thermal Melt Spectra Depicts the Structural Stability of the Mutated Proteins.

The stability of the mutant protein was measured through thermal denaturation since stably folded proteins tend to display a sigmoidal shape. This distinct sigmoid shape is due to the temperature dependent equilibrium between two states, the folded state and the unfolded state. The thermal melt data was fit with a curve based on the two-state model for the unfolding equilibrium in order obtain the melting temperature (T_m) values. Many proteins aggregate or precipitate quickly after they are unfolded which makes unfolding irreversible. The reversibility of the unfolding reaction can be assessed by cooling the sample to see if the folding reaction is duplicated. If and only if the melting is fully reversible, the melting temperature is directly related to conformational stability, and the thermodynamics of protein folding can be extracted from the data (Pace et al. 1996). Thermal melt experiments on Pfl 6 (Figure 10) and Xfaso 1 (Figure 11) mutants were performed at 222 nm to determine whether the distinct mutations had a stability comparable to their respective wild types.

Mutations in Pfl 6

The Leu6Ile mutation in Pfl 6 was seen as only slightly destabilizing seen by the small decrease in the T_m value, from a Pfl 6-wt of 64°C to the experimental mutant value of 62°C, when the protein underwent denaturation. Also, the forward and reverse thermal melts show the reversibility of the protein denaturation. The Pfl 6 mutant Tyr9Ala had a T_m value of 50°C and the melts were reversible. The His13Leu mutation had an elevation in the T_m

value to 70°C with both the forward and reverse melts indicating reversibility. Lastly, the triple mutations Leu6Ile/Tyr9Ala/His13Leu in Pfl 6 had a reversible T_m value of 50°C.

Mutations in Xfaso 1

The variant Ile4Leu had a reversible thermal melt T_m of 52°C which shows that this variant is close to the folded wild type stability by comparing the wild type T_m of 55°C. Additionally the denaturation melt curve for Ala7Tyr variant in Xfaso 1 does not have much of a melt nor very much folded structure. Compared to the other variants the Ala7Tyr variant is largely unfolded. The thermal melt for Leu11His coincided with the far-UV data by producing a lower T_m of 50°C than the wild type of 55°C which indicates the slight destabilization and a possible shift towards the unfolded state of the protein.

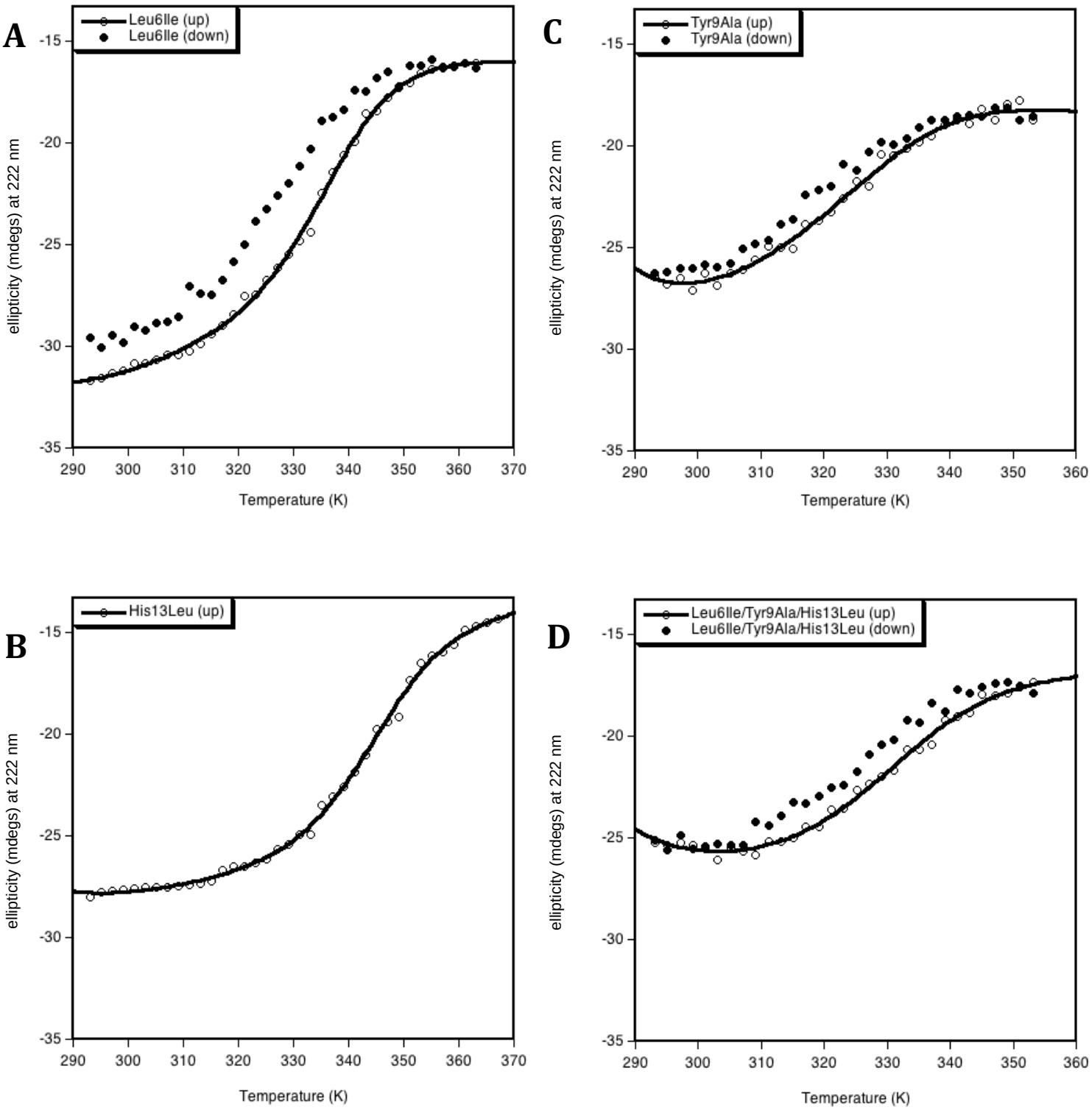

Figure 10: Forward (up) and reverse (down) Thermal Denaturation Curves for Pfl 6 mutants: (A) Leu6Ile, (B) His13Leu, (C)Tyr9Ala and (D)Leu6Ile/Tyr9Ala/His13Leu were all taken from 20-80°C except for (B) His13Leu which had a max temperature of 94°C. The thermal melt data were collected in 2 mm pathlength cells at 222 nm. Note that (B) does not have a reverse melt that proceeds from 367 K to 293 K, but other forward and reverse melts over narrower temperature ranges of His13Leu confirmed reversibility.

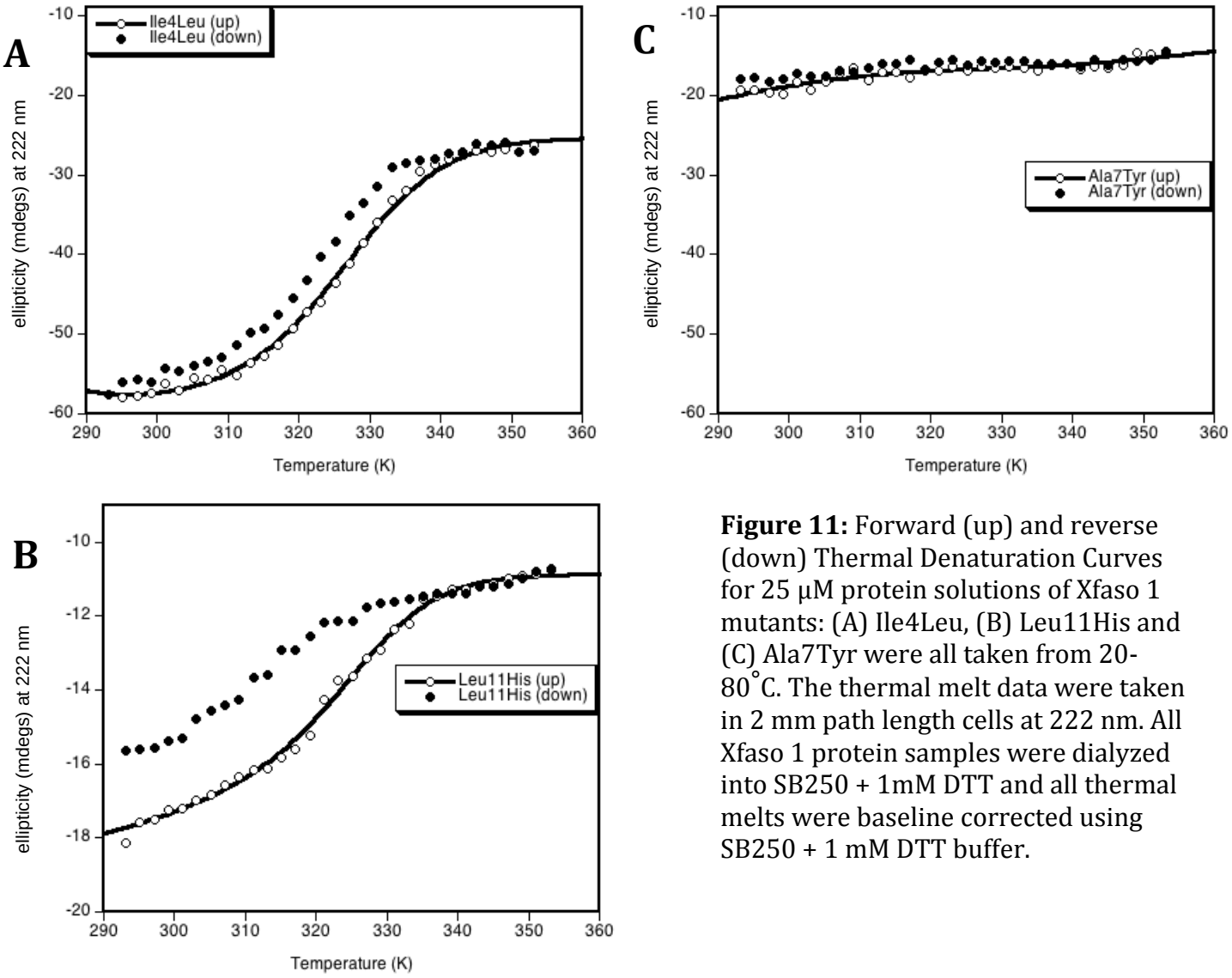

Figure 11: Forward (up) and reverse (down) Thermal Denaturation Curves for 25 µM protein solutions of Xfaso 1 mutants: (A) Ile4Leu, (B) Leu11His and (C) Ala7Tyr were all taken from 20-80˚C. The thermal melt data were taken in 2 mm path length cells at 222 nm. All Xfaso 1 protein samples were dialyzed into SB250 + 1mM DTT and all thermal melts were baseline corrected using SB250 + 1 mM DTT buffer.

Summary of Results

Table 1: Summary of Data from far-UV spectroscopy and Thermal Melt Denaturation curves.

Variant	208/222 ratio	T_m (K)	T_m (°C)	Assessment relative to WT
Pfl 6-wt	1.24	337	64	NA
Pfl 6-Leu6Ile	1.14	335	62	slightly destabilized
Pfl 6-Tyr9Ala	1.6	323	50	significantly destabilized
Pfl 6-His13Leu	1.26	343	70	slightly stabilized
Pfl 6-Leu6Ile/ Tyr9Ala/ His13Leu	1.54	324	50	significantly destabilized
Xfaso 1-wt	1.24	328	51	NA
Xfaso 1-Ile4Leu	1.16	325	52	slightly destabilized
Xfaso 1-Ala7Tyr	2.00	311	< 30	significantly destabilized
Xfaso 1-Leu11His	1.79	323?	50?	significantly destabilized?

Discussion

In this study, we examined which features of the Xfaso 1 and Pfl 6 sequence are critical for the fold each protein adopts, to illuminate possible mechanisms leading to structural change.

The variant in Pfl 6 Leu6Ile is interpreted by the far-UV data and the thermal melt as slightly destabilized, with the α-helical structure indicator of the 208/222 ratio being less than the Pfl 6 wild type ratio which may indicate a rearrangement of α-helices (Steinmetz et al. 1998). The Pfl 6 mutant Tyr9Ala was seen to be significantly destabilized from the large decrease in the T_m value and the increase in 208/222 ratio. The data indicates that it is probable that the Pfl 6 mutant Tyr9Ala most likely resided partly in the unfolded protein state. A large factor that could play a role in this shift is the characteristics of the amino acid switch. The large-sized polar tyrosine group was swapped for a smaller-sized nonpolar alanine group which most likely caused a cavity in the interfaces of the protein. For the Pfl 6 protein, the Tyr9Ala change could have been too drastic to maintain the folded state of the protein. In addition, the His13Leu mutation in Pfl 6 is slightly stabilizing perhaps because of introduction of the less polar Leu group into the interior of protein. His13Leu is slightly stabilizing because of the less polar Leu group in the interior of protein. This hydrophobic comparison between His and Leu serves to explain the increase in stability of the protein while maintaining the α-helical structure. It was not initially anticipated for this mutation to have an increase in T_m value and new thermal melts were measured at a broader and higher temperature range to 94°C in order to adequately measure the unfolded baseline. Then, all three of the mutations were inserted into Pfl 6 in anticipation

to see if the effects of the three mutations were additively contributing to the unfolded state of the protein. As anticipated, the Pfl 6 triple mutant was interpreted as significantly destabilized, but the values of the 208/222 ratio and the T_m value were close that of Pfl 6 Tyr9Ala mutation. This observation probably shows the importance of the residue at the 9 position of Pfl 6 and it exerts a more dominant role in the protein's secondary structure.

The reversal of these mutations was performed in Xfaso 1. In the Xfaso 1 mutant Ile4Leu, there was a decrease in the 208/222 ratio and this decrease was common to both Xfaso 1 and Pfl 6 proteins. The T_m value was also slightly higher in comparison to the wild type T_m value which indicates stability of this mutation and that this position in the protein tolerates the mutation. Then the Xfaso 1 mutant Leu11His was analyzed to have a large 208/222 ratio which probably indicates the transition of the protein to an unfolded state. Along with the thermal melt data, Leu11His seemed perplexing because the thermal melt did not seem as reversible as the other mutants and with this factor , the T_m value was strangely close to the Xfaso 1 wild type T_m value. Based on the information from the Pfl 6 mutant, the Leu to His switch was expected to destabilize the protein because the hydrophobic nature of this region in the two proteins. The data from Xfaso 1 complicates the interpretation of the toleration of this mutation since the T_m value shows the stability of the mutant protein, whereas the far-UV data shows the loss of α-helical structure. The third mutation made in Xfaso 1, Ala7Tyr, correlated with its mutant counterpart in Pfl 6. The mutation was perceived as causing the protein to reside in an unfolded state with loss of α-helical structure. The two most apparently destabilizing mutations are the ones that change from Ala to Tyr or vice versa. This goes along with idea of the substitution of larger and

more polar amino acid which causes unfavorable interaction between helix 1 and the divergent C-terminal end of the two respective proteins.

The reversals of the mutation seen in both Pfl 6 and Xfaso 1 show three different cases of the interface between helix 1 and the C-terminus. For the swaps between Ile and Leu, the location and the amino acid did not destabilize the protein nor disrupt the α-helical structure. The swaps between Tyr and Ala, showed that Xfaso 1 and Pfl 6 preferred their respective amino acid residues in order to maintain the stability of the protein. Lastly, both Xfaso 1 and Pfl 6 better tolerated the Leu as opposed to the His amino acid.

It is likely that the structural changes of Xfaso 1 and Pfl 6 occurred through multiple and specific changes in their sequences, in which the order that these events occurred plays an important role. Therefore at this time it is not enough evidence to make connections of the mechanism by which Xfaso 1 evolved from the all α-helical structure to the α+β folded structure of Pfl 6. Moreover additional studies have shown that without large-scale mutations, new protein folds can evolve from existing folds (Cordes et al., 1999). In this study they showed that the simple switch of the sequence positions of two adjacent residues in a β-sheet of the arc repressor homodimer, the β-sheet switched to a pair of 3_{10}-helices (Cordes et al., 2003). This study shows promise that structural transformations can actually occur with minimal change in protein sequences.

The next direction to go in for this project involves the stability of the histidine to leucine transition found in Pfl 6. Pfl 6 is pH dependent due to the histidine residue. At lower pH Pfl 6 is destabilizing, but the mutant His13Leu should not be destabilizing from the pH factors

due to the lack of the imidazole group. By this approach, the capability of the leucine to stabilize the protein should be maintained and provide another form of analysis of protein evolution. The ideal goal for this project is to create and to support a theory on the mechanism of how proteins can shift between an all α-helical structure to an α+β folded protein.

Acknowledgements

I would like to thank Dr. Cordes for giving me the opportunity to work in his lab for the past year. Additionally, I would like to thank Dr. Cordes for his crucial role in the development and analysis of this project and making my research experience a captivating adventure. I would like to thank the Cordes lab for making my research experience at the University of Arizona pleasant and enjoyable. The direction and patience of Katie Holso was much appreciated throughout this project. Thank you to Dr. Chad K. Park for advice on circular dichroism.

This research was supported by NIH grant R01 GM066806 to M.H.J.C

Literature Cited

[1] Allen R. J., Warren, P. B. and ten Wolde, P. R. (2005) Sampling rare switching events in biochemical networks, *Physical Review Letters* 94, 018104: 1-4.

[2] Cordes, M.H., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (1999) Evolution of Protein Fold in Vitro. *Science* 284, 325-328.

[3] Edelhoch, H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* 6, 1948-1954.

[4] Gill, S.C., von Hippel, P.H. (1989) Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem*. 182, 319-326.

[5] Grishin, N.V. (2001) Fold Change in Evolution of Protein Structures. *Journal of Structural Biology* 134, 167-185.

[6] "Life After the Lac Operons." Biology Hypertextbook Index. Science Gateway, 01 Jan 2008. Web. 13 Apr 2011. <http://www.sciencegateway.org/resources/biologytext/pge/pgeother.html>.

[7] Minor, D.L. Jr., Kim, P.S. (1996) Context-Dependent Secondary Structure Formation of a Designed Protein Sequence. *Nature* 380, 730-734.

[8] Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* 247, 536-540.

[9] Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995) Denaturant m values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci*. 10, 2138-2148.

[10] Newlove, T., Konieczka, J.H., and Cordes, M.H.J. (2004) Secondary Structure Switching in Cro Protein Evolution. *Structure* 12, 569-581.

[11] Newlove, T., K.R. Atkinson, L.O. Van Dorn, and M.H.J. Cordes. (2006) A Trade Between Similar but Nonequivalent Intrasubunit and Intersubunit Contacts in Cro Dimer Evolution. *Biochemistry.* 45, 6379–6391 (2006).

[12] Pace, C.N. and J.M. Scholtz, (1996) Measuring the conformational stability of a protein in "Protein Structure: A Practical Approach" (Creighton, T. E.) pp. 299-321, IRL Press, Oxford.

[13] Steinmetz, M.O., Stock, A., Schulthess, T., Landwehr, R., Lustig, A., Faix, J., Gerisch, G., Aebi, U., and Kammerer, R.A. (1998). A distinct 14-residue site triggers coiled-coil formation in cortexillin I. EMBO J. 17, 1883-1891.

[14] Van Dorn, L.O., Newloe, T., Chang, S., Ingram, W.M., and Cordes, M.H.J. (2006) Relationship between Sequence Determinants of Stability for Two Natural Homologous Proteins with Different Folds. *Biochemistry* 45, 10542-10553.

Figure Legends

Figure 1: Diversity of Protein folds. A limited depiction of structural topologies of proteins that encompass a broad diversity (Courtesy of Matthew H.J. Cordes).

Figure 2: The Lambda Phage Cycle Decision. CI starts to win the competition when the bacteria are not growing very well, because it is continuously being made at a low level. When the bacteria are not dividing as often, the concentration builds up and CI inhibits the lytic phase (Adapted from "Biology Hypertextbook Index").

Figure 3: Transitive Homology Analysis. Three prophage Cro proteins, Afe01, Xfaso 1, and Pfl 6 were used as sequence intermediates in a transitive homology analysis connecting P22 Cro to λ Cro. P22 Cro is compromised of an all α-helical fold, while λ Cro is comprised of a mixed α+β fold. Xfaso 1, Pfl 6, and Afe01 are intermediate in sequence to P22 Cro and λ Cro. Percent sequence identities (%ID) are given for each link in the chain (adapted from Roessler et al. 2008).

Figure 4: Pfl 6 and Xfaso 1. Homologous proteins with 40% sequence identity and different folds (Roessler et al. 2008).

Figure 5: Interactions between the N-terminus and the C-terminal structures of Pfl 6 and Xfaso 1. The C-terminus is in gold and the N-terminus in purple or green. The three major packing side chains from Pfl 6 are Leu 6, Tyr 9 and His 13 and for Xfaso 1 they are Ile 4, Ala 7, Leu 11. Interacting side chains from the C-terminus are shown in dotted spheres (Courtesy of Matthew H.J. Cordes).

Figure 6: Monitoring with gel electrophoresis of (A) Pfl 6 mutagenesis of Tyr9Ala mutation and (B) Xfaso 1 mutagenesis of Ala7Tyr. The 1 kb ladder marker of 6,108 kb is noted.

Figure 7: Coomassie-stained gel (15% SDS-PAGE) of the four purified Pfl 6 variant eluates by Ni-affinity chromatography from dialysis were seen to be pure via SDS-PAGE Tris-Tricine gel electrophoresis.

Figure 8: Far UV Spectra of Pfl 6 Mutants: 25 μM mutant Leu6Ile, Tyr9Ala, and His13Leu and the triple mutant were scanned at 20˚C from 205-260 nm in a 1 mm pathlength cell. All Pfl 6 protein samples were dialyzed into SB250 and spectra were both baseline-corrected with SB250.

Figure 9: Far UV Spectra of Xfaso 1 Mutants: 25 μM mutant of each mutant Ile4Leu, Ala7Tyr, and Leu11His were scanned at 20˚C from 205-260 nm in a 1 mm pathlength cell. All Xfaso 1 protein samples were dialyzed into SB250 + 1 mM DTT and spectra were both baseline-corrected with SB250 + 1 mM DTT.

Figure 10: Forward (up) and reverse (down) Thermal Denaturation Curves for Pfl 6 mutants: (A) Leu6Ile, (B) His13Leu, (C)Tyr9Ala and (D)Leu6Ile/Tyr9Ala/His13Leu were all taken from 20-80˚C except for (B) His13Leu which had a max temperature of 94˚C. The thermal melt data were collected in 2 mm pathlength cells at 222 nm. Note that (B) does not have a reverse melt that proceeds from 367 K to 293 K, but other forward and reverse melts over narrower temperature ranges of His13Leu confirmed reversibility.

Figure 11: Forward (up) and reverse (down) Thermal Denaturation Curves for 25 µM protein solutions of Xfaso 1 mutants: (A) Ile4Leu, (B) Leu11His and (C) Ala7Tyr were all taken from 20-80˚C. The thermal melt data were taken in 2 mm path length cells at 222 nm. All Xfaso 1 protein samples were dialyzed into SB250 + 1mM DTT and all thermal melts were baseline corrected using SB250 + 1 mM DTT buffer.

Table 1: Summary of Data from far-UV spectroscopy and Thermal Melt Denaturation curves.