THE ROLE OF BURIED STACK RESIDUES IN THE FOLDING OF THE BETA-HELIX DOMAIN OF P22 TAILSPIKE

by

Ryan Matthew Simkovsky

B.S. Biology
California Institute of Technology, 2001

Submitted to the Department of Biology in partial fulfillment
of the requirement for the degree of

DOCTOR OF PHILOSOPHY
in Biology

at the
Massachusetts Institute of Technology

June 2007

Signature of Author_____
Department of Biology
June 2007

Certified by_____
Jonathan King
Thesis Supervisor

Accepted by_____
Stephen P. Bell
Chairman, Department Committee on Graduate Students

THE ROLE OF BURIED STACK RESIDUES IN THE FOLDING OF THE BETA-HELIX DOMAIN OF P22 TAILSPIKE

By Ryan Matthew Simkovsky

Submitted to the Department of Biology at the Massachusetts Institute of Technology on June 26, 2007 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology

ABSTRACT

In-register, parallel alignment of similar or identical side chains is a common structural phenomenon in amyloid fibers. In the crystal structure of an amyloid fiber, these residues not only align, but take identical side chain orientations, creating long stacks of identical residues with identical orientations. This phenomenon of side chain stacking is a dominant component of the amyloid structure, and may take an equally important role in the seemingly sequence-independent formation of amyloid fibers or the sequence-dependent species barrier to amyloid transmission.

Side chain stacking has long been observed as a prominent feature of the soluble parallel β-helix fold, which exhibits a cross-β structure reminiscent of amyloids. To investigate the sequence requirements of these stacks for directing a polypeptide chain into a β-helical fold, we performed systematic mutational studies using the complete P22 tailspike protein, which contains a 13 rung β-helix domain. The *in vivo* folding and assembly of 150 single mutant polypeptide chains were characterized at multiple temperatures by SDS-PAGE, which distinguishes the fully native, SDS-resistant state from partially folded or misfolded conformers. The vast majority of the buried core was completely intolerent to subsitution at physiological temperatures, while only a few sites composing a continuously contacting network of residues called the folding spine were intolerant of alanine mutations at 18ºC.

These results indicate that the β-helix folds through a processive folding mechanism, analogous to the nucleation and elongation mechanism of amyloids, but requiring recurring sequence-dependent stacking contacts that stretch the length of the β-helix.

Numerous avenues of investigation demonstrated that the observed folding defects of the full length tailspike were due to a sequence effect on the folding of the β-helix domain. Alanine mutants capable of reaching a native-like trimeric state were as thermostable as the wild-type trimer, indicating that mutants affect a folding intermediate rather than the native state. *In vivo* folding efficiencies of eight β-helix folding mutants in the context of the isolated β-helix domain paralleled the folding of the full length tailspike, demonstrating that these mutants act at the level of β-helix folding. *In vitro* refolding of both full length and isolated β-helix tailspike mutants demonstrated that the observed mutant folding deficiencies were solely caused by the single amino acid change and not due to interactions with cellular components.

Folding spine mutations were further tested in the context of a suppressor mutation, A334V, which was previously shown to stabilize the β-helix and rescue classic temperature sensitive folding mutations. Most mutations could not be rescued by the addition of A334V, a number of which were in fact further hindered in folding by the presence of the suppressor mutation. Eight folding spine mutations located between rungs 4 and 6 were rescued in a position specific manner, though only one of these positions directly contacts the suppressor. The pattern of rescue and hinderance observed with double mutants indicates that A334V is not a global suppressor and instead acts on a likely nucleus for folding.

These studies provide evidence for a nucleation and elongation mechanism for folding of the parallel β-helix. Single mutations throughout the β-helix are capable of blocking this step-wise folding mechanism in a temperature dependent manner that likely leads to the population of a partially folded β-helix species that would be inherently aggregation prone.

Thesis Supervisor: Jonathan King, Professor of Biology

ACKNOWLEDGEMENTS

There are so many people who have helped me through the past five years, that in no way can this list of acknowledgements appropriately thank everyone. Thank you to everyone whose paths have intertwined with mine over these past years and who have made my life a happy one.

Professor Jonathan King, you have been an inspiration, a mentor, a leader, a supporter, a listener, a helper, and a guide. I cannot thank you enough for opening your lab to me, for all of the time you spent in teaching me to hone my mind and skills as a scientist and as a person. Thank you for your patience and understanding. Thank you for balancing my independence with the right amount of guidance and intervention. I appreciate to no end your abilities to train me as an individual. I thank you and your family for letting me share in your life. You have been an ideal mentor through these past years, and I will never be able to truly express my gratitude.

Cameron Haase-Pettingell, you are the mother of the lab, the one who keeps the whole thing running. You are also the local expert on all of the tailspike experiments that never quite made it into the literature, and that has proven invaluable. You taught me how to run gels, which proved to be the most frequently used technique in my research. Every time I turn around, you were trying your best to help out, even if it was something as simple as volunteering to make me a gel. Thank you for all of your help, support, expertise, and knowledge.

Dr. Peter Weigele, you have been both a colleague and a friend. Scientifically, you gave me a jump start. You were instrumental in getting the high throughput expression system going. You constantly challenged me and help me grow through our daily discussions. Personally, you've been the one to make the lab into the fun place that I and others enjoy coming to for work. Of course, it has also been great playing hooky with you, whether it was simply for a meal or for a canoe trip up river.

To the rest of the King lab, past and present, I thank you all for making this a second home, a wonderful place to work and to come in every day. All of the graduate students who came before me and have now since graduated, Dr. Melissa Kosinski-Collins, Dr. Welkin Pope, Dr. Shannon Flaugh, and Dr. Ishara Mills, were constant sources of advice and experience whenever I faced a new challenge. My other concurrent graduate students, Jiejin Chen, Ligia Acosta, Greg Pintilie, Desislava Raytcheva, and Kate Drahos have given technical help over the years, collaborated with me on new projects such as computer programs or trying to get a new machine working, helped proof read this thesis, and have simply added to the upbeat atmosphere of the lab. Thank you to Dr. Yongting Wang, who was a wonderful person to bounce ideas off of and an inspiration for many new experiments. Thank you for being such a patient ear. Thank you Dr. Kelly Knee for your input on the analysis of my kinetic data.

I want to specially thank Kristen Cook, a past undergraduate in the lab who helped with some of the initial mutagenesis reactions; Phillip Campbell, a past undergraduate who helped construct the His-tagged full-length constructs; and Althea Hill, a high school student who bravely took on the challenge of drying down hundreds of my triple-wide gels. Thank you to Professors David Gossard and Jacqueline Piret, who gave invaluable feed back, advice, and training. A special thanks to Cindy Woolley, without whom many deadlines and crossword puzzles would not have been made

possible. And thank you to Margaret, who often kept me company on those late nights in lab and sometimes chased me out of the lab so that I could get some sleep.

I must of course thank my thesis committee, Susan Lindquist, Amy Keating, David Housman, and Robert Seckler. Not only were many of you instrumental in providing me with advice and guidance, but I have been lucky to be able to collaborate with nearly all of you at one point in time, such as my work with James Apgar in the Keating lab to compute energies associated with my mutants and Dr. Peter Tessier in the Lindquist lab to see if native tailspike bind thioflavin T (it did not in my hands). I especially want to thank Dr. Robert Seckler and his graduate student, Anait Seul, who showed me great hospitality and generosity in welcoming me to their lab in Potsdam, Germany, in teaching me everything I know about how to work with the isolated β-helix domain, for providing me with a β-helix-encoding vector, and for showing me around Potsdam and Berlin.

I would like to thank Dr. Sean Decatur and the entire Decatur lab. You have been close colleagues and collaborators, and I have enjoyed my visits to your lab and our discussions through the years. At every biophysical society meeting, it became my pleasure to become an honorary Decatur lab member, and I can't thank you all enough for that companionship.

Thank you to the Horvitz, Keating, Sinskey, and Baker labs for the use of their equipment. Also, the mass spectrometry work discussed in Appendix B was performed by members of the MIT Biolpolymers facility, Richard Cook, All Leshinsky, and Katie Tone. Thank you to Dr. Anne Robinson, who provided me with the pET(Δ*N*) vector.

For the past four years, I have not only been in lab doing work, but I have also been a graduate resident tutor (GRT) for Third East in East Campus, also known as Tetazoo. Over 80 students have touched my life from that hall, and I can't thank you enough for keeping me insane, providing me with plenty of fun distractions and new challenges through the years. My life at MIT would definitely have never been complete without my experience and time on the hall, and I hope that I have given as much to all of you that you have given to me.

Thank you to all of my friends and classmates who were there with me through the years. Although this list is certainly too numerous to give here, I want to especially thank my classmate, colleague, and friend Emiko Bare, who not only helped me with protein purification in her lab, but who was instrumental in getting me back to Hawaii.

I want to thank the entire Hanovice family, Ron, Cathryn, Gordon, and Nick for their love and support. I especially want to thank Ron for helping me study for my prelims and always challenging my teaching abilities with large stacks of medical journals. Thank you to Nick for helping me out in lab on a couple of occasions and for also giving me the opportunity to teach a little biology.

I cannot thank my own family enough for their love and support for the past 27 years. Without you, I would certainly not be writing these thanks, and for that I am eternally grateful. I love you all very much.

And last, but certainly not least, I must thank my special lady friend, Rebecca Hanovice, Esq. Without your continual love over these past four years, I don't think I would ever have made it through.

This research was supported by NIH GM17980 awarded to Jonathan King.

BIOGRAPHICAL NOTE
Ryan Matthew Simkovsky

Education

Ph.D. Expected June 2007	Massachusetts Institute of Technology (MIT) Department of Biology, Cambridge, MA.
B.S. June 2001	California Institute of Technology (Caltech) Department of Biology, Pasadena, CA.

Research and Professional Experience

2003 to 2007	Graduate Research Assistant in the laboratory of Professor Jonathan King, MIT Department of Biology, Cambridge, MA.
2001 to 2002	Computational Biology Programming Lead at Arbor Vita Corporation, Sunnyvale, CA.
2000 to 2001	Undergraduate Research Assistant in the laboratory of Professor Francis Arnold, Caltech Department of Chemical Engineering, Pasadena, CA.
2000 to 2001	Undergraduate Research Assistant in the laboratory of Professor Niles Pierce, Caltech Department of Applied and Computational Mathematics, Pasadena, CA.
1999 to 2000	Undergraduate Research Assistant in the laboratory of Professor Ken Nealson, Caltech Department of Geology, Pasadena, CA.
Summer 1999	Undergraduate Research Fellow in the laboratory of Professor Ken Nealson, Jet Propulsion Laboratory, Astrobiology Division, Pasadena, CA.
Summer 1998	Research Intern at Canji, Inc., La Jolla, CA.
Summer 1996	Research Intern in the laboratory of Dr. Howard Fox, The Scripps Research Institute, Neuropharmacology Division, La Jolla, CA.

Research and Professional Experience

Journal Articles

Simkovsky, R. and King, J. (2006) "An elongated spine of buried core residues necessary for *in vivo* folding of the parallel β-helix of P22 tailspike adhesin". *Proc Natl Acad Sci USA*, 103, 3575-3580.

Conference Papers

King, J., Haase-Pettingell, C., Simkovsky, R., Weigele, P. "Amino Acid Sequence Control of the Folding of the Parallel β-Helix, the Simplest β-Sheet Fold." *Lecture Notes in Computer Science: RECOMB* (2005): 472-473.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Aβ	Amyloid-β protein
ATP	Adenosine triphosphate
βHX	Isolated β-helix domain protein
BME	β-mercaptoethanol
CASP	Critical Assessment of Structure Predication
CD	Circular dichroism
CFTR	Cystic fibrosis transmembrane conductance regulator
CRABPI	Cellular retinoic acid binding protein I
DMSO	Dimethyl sulfoxide
ΔG	Gibbs free energy of folding at 0 M denaturant
ΔN	Tailspike fragment protein containing residues 109-666
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
IPTG	Isopropyl-β-D-thiogalactoside
I_U	Tailspike trimeric unfolding intermediate
LB	Luria-Bertani broth
$MgSO_4$	Magnesium sulfate
MOPS	3-[N-morpholino]-propanesulfonic acid
NaCl	Sodium chloride
NAH_2PO_4	Sodium phosphate
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NT	P22 Tailspike native trimer
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PMSF	Phenylmethanesulphonyl flouride
RP	Retinitis pigmentosa
SDS	Sodium dodecyl sulfate
su	Suppressor of *tsf*
TRIS	Tris(hydroxymethyl)-amino methane
tsf	Temperature sensitive for folding
UV	Ultraviolet
WT	wild-type

CHAPTER ONE:

INTRODUCTION

A. PROTEIN MISFOLDING IS A MEDICAL PROBLEM

In order for the human body to work properly, cells must not only synthesize proteins but must ensure that these proteins fold into the structure that allows them to perform their evolved task. The primary control of a protein's ability to fold into a specific, native three-dimensional structure is the sequence of amino acids in the polypeptide chain itself. In 1961, Anfinsen demonstrated that a mammalian protein, RNase, could refold spontaneously without the aid of cellular factors (Anfinsen et al., 1961; Haber and Anfinsen, 1961).

Although RNase is capable of spontaneously folding in the test tube, Anfinsen and Haber demonstrated that folding yields dropped off significantly as the protein concentration increased to physiologically relevant levels (Anfinsen and Haber, 1961). Subsequent research revealed that a number of proteins necessary for the proper function of the cell cannot fold from sequence alone and require the activity of chaperone proteins to reach their native state. For example, the structural protein actin requires the presence of a specific chaperonin, TriC, in order to fold (Gao et al., 1993). Environmental stresses, such as increases in temperature, enhance the incidence of protein misfolding. In response, most tissues increase their production of chaperone proteins, particulary a group of chaperones known as the heat shock proteins (Whitesell and Lindquist, 2005). The failure of purified proteins to refold *in vitro* can be due to the absence of a required chaperone, other essential components, or to intrinsic limitations in the sequence control of folding.

When proteins are unable to fold to their native state, they cannot perform their functions. The "natively unfolded" proteins may be an exception to this (Fink, 2005). In the case of the cystic fibrosis transmembrane conductance regulator (CFTR), the most common disease-associated mutation, ΔF508, prevents the protein from folding into its native conformation in the cell (Qu et al., 1997). The misfolded protein is retained in the endoplasmic reticulum and is quickly degraded before it can be trafficked to the surface of epithelial cells (Rowe et al., 2005). The lack of folded CFTR proteins on the cell's surface leads to improper regulation of osmosis, which in turns leads to the symptoms associated with cystic fibrosis.

Although loss of function through an inability to fold can lead to disease, it has become apparent in the past decade that misfolded proteins may go on to aggregate. Aggregation can disrupt the homeostasis of the cell and lead to cell death and disease (Dobson, 2006). Although proteins may precipitate out of solution in a non-specific manner that depends on environmental conditions and protein concentrations, a number of protein misfolding and aggregation diseases have been discovered that depend upon the protein's amino acid sequence. Diseases involving aggregation or misfolding of polypeptide chains are becoming increasingly common, particularly as the average human life span increases in the modern world. In order to develop therapeutics towards these classes of diseases, it is important to understand how polypeptide chains misfold or aggregate and how the amino acid sequence of these proteins contribute to the development of these diseases.

1. Native-state Polymerization: Sickle Cell Hemoglobin

Sickle cell anemia is a disease in which mutant hemoglobin aggregate through the interactions of natively folded proteins. A single amino acid change, a glutamine to valine mutation at position six of β-hemoglobin, causes native hemoglobin molecules to polymerize (Ingram, 1956). This substitution of a charged surface residue with a small hydrophobic group allows the association of hemoglobin molecules through the interaction of the valine mutant residue with a surface hydrophobic pocket comprised of F85 and L88 on the β-hemoglobin subunit of another molecule (Bihoreau et al., 1992). This interaction can only occur under low oxygen conditions due to the fact that the hydrophobic pocket is buried in the oxygenated state and only reveals itself on the surface of β-hemoglobin upon a conformational change induced by deoxygenation (Wishner et al., 1975). Since every hemoglobin tetramer contains two β-hemoglobin subunits, repetition of these sickling interactions with more hemoglobin molecules generates long polymerized fiber-like threads.

As stated, low oxygen conditions induce polymerization across the erythrocyte, causing the red blood cell to stiffen and stretch into its sickle shape. This polymerization process proceeds through a double nucleation and polymerization mechanism (Ferrone et

al., 1985). In this mechanism, the first phase of aggregation is a homogenous polymerization that is induced by the formation of a slow-forming and metastable oligomer through the hydrophobic binding described above. Single unbranched fibers can then proceed to polymerize from this nucleus. In the second heterogeneous nucleation, fibers are seeded on the non-growing surface of pre-existing fibers, allowing for the generation of thicker, branched fibrils and gels.

2. Loop-Sheet Insertion: The Serpins

Serine protease inhibitors, or serpins, are the largest and most widely distributed superfamily of proteins that regulate proteolytic cascades (Irving et al., 2000). This is most typified by α_1-antitrypsin, which controls proteases that are involved in inflammatory cascades. This superfamily also includes anti-thrombin, C1 inhibitor, α_1-antichymotrypsin, and plasminogen activator inhibitor-1. These latter proteins play important roles in the control of coagulation, inflammation, as well as other protease cascades (Lomas et al., 2005). Although quite functionally diverse, the superfamily has a high degree of sequence similarity and a conserved tertiary structure. This conserved structure is composed of three β-sheets and an exposed mobile loop.

Inhibition of proteases occurs through a mousetrap-like system. The exposed mobile loop contains a peptide sequence that is a pseudosubstrate for its target protease. Critical residues (P1-P1') in this loop act as bait for the protease, which cleaves the peptide bond between these two residues (Johnson and Travis, 1978; Wilczynska et al., 1995). Upon cleavage, the mobile loop of the serpin swings the covalently linked protease over the entire length of the serpin, approximately 70 Å, while the reactive loop inserts itself into β-sheet A to form an additional β-strand (Huntington et al., 2000). This drastic change leaves the protease in a deformed and irreversibly inactive conformation. The resulting conformation of the serpin-protease complex is subsequently recognized for degradation. The large shift in conformation of the serpin is possible due to the fact that the serpins exist in a kinetically trapped, metastable active form. Upon peptide bond cleavage, the kinetic trap is released and the final loop-sheet insertion conformation is more thermodynamically stable that the initial structure (Carrell and Gooptu, 1998).

Point mutations in the reactive loop or the β-sheet region of the serpins can lead to distortion of the structure that allows for the insertion of the reactive loop of one molecule to lock into the β-sheet of a second molecule to form a dimer. This loop-sheet insertion oligomerization can ultimately lead to polymerization of serpins. In the case of α_1-antitrypsin, the Z mutation (E342K) generates a structural distortion that allows for polymerization. These polymers may either be degraded (Wu et al., 2003) or accumulate in inclusion bodies in the endoplasmic reticulum (Janciauskiene et al., 2004). Alternatively, the mutation may lead to an early insertion of the loop into the β-sheet, generating an inactive latent confirmation (Silverman et al., 2001). Seeding has been shown to accelerate the polymerization of serpin proteins, indicating that a nucleation dependent process may describe this specific aggregation process (Crowther et al., 2003).

Polymerization of the serpins leads to two medically relevant outcomes, one being the accumulation of the polymers in liver tissue and the second being the failure to secrete serpins that leads to plasma deficiency. The accumulation of misfolded α_1-antitrypsin in liver cells leads to the development of cirrhosis (Lomas et al., 2005). Neuroserpin polymerization and accumulation in neurons causes dementia (Miranda and Lomas, 2006). Deficiency of a number of different serpins can lead to clinical syndromes such as thrombosis, angioedema, and emphysema.

3. Domain Swapping

The loop-sheet insertion mechanism of the serpin polymerization reactions is an example of a broader class of oligomerization mechanisms known as 3D domain swapping. This general mechanism of protein oligomerization occurs when two or more proteins swap identical structural features, such as secondary structural elements, with each other to form dimers or oligomers (Liu and Eisenberg, 2002). This results in native-like intermolecular contacts being formed in place of their intramolecular counterparts. Exchanged structural elements range in size and complexity. In crystal structures of domain swapped dimers of RNase A, the swapped components were single units of secondary structure, such as a single α-helix or a single β-strand (Liu et al., 2001; Liu et al., 1998). In contrast, entire domains constituting half of the protein may swap, as is

apparent when the native structures of the dimeric βB2-crystallin and the homologous monomeric γB-crystallin protein are compared (Bax et al., 1990).

Domain swapped proteins may exist in closed conformations, where all domains are partnered with another protein. Dimers and trimers have been crystallographically observed for RNase A (Liu et al., 2001; Liu et al., 2002; Liu et al., 1998), which have led to proposed models for tetramers and the notion of higher order domain-swapped oligomers (Liu and Eisenberg, 2002). Domain swapping may provide a functional advantage, as is the case for glyoxalase I where the domain swapped homodimer displays enhanced enzymatic activity in comparison to the monomeric state (Saint-Jean et al., 1998).

While closed domain swapping may result in functional proteins, aggregation may be a result of an open conformation of domain swapping where swapping may produce unsatisfied domains that can cascade into a monomer addition aggregation pathway (Rousseau et al., 2003). Kinetic studies on domain swapping have revealed that the mechanism of domain swapping is dependent upon the properties of the protein and that there is not a universal mechanism for domain swapping. RNase domain swapping appears to occur through a partially folded monomeric intermediate (Liu et al., 2001). In contrast, the cell cycle regulatory protein p13sucI requires complete unfolding of its monomer in order to generate a domain-swapped dimer from the unfolded state (Rousseau et al., 2001). The kinetics of domain swapping of p13sucI correlated with the rates of aggregation upon heating, indicating that aggregation may occur through a domain swapping mechanism (Rousseau et al., 2001).

Domain swapping has been proposed as a mechanism for the aggregation of the human γD-crystallin protein (Flaugh et al., 2006). This aggregation is likely a strong contributor to the eye disease cataract, which is the leading cause of blindness in the world. The γD-crystallin protein is composed of two domains that display differential stability, with the C-terminal domain being more stable than the N-terminal domain (Flaugh et al., 2005a). Mutants in the interface between these domains destabilize the N-terminal domain resulting in a proposed unfolding intermediate where the C-terminal

domain is folded but the rest of the protein is unfolded (Flaugh et al., 2005b). Since C-terminal domain interface residues nucleate the folding of the N-terminus, aggregation of the protein may proceed through a domain swapping mechanism where the C-terminus of one monomer would refold the N-terminus of another monomer, which then in turn acts on another monomer resulting in linear polymerization. Oxidative damage can cause deamidation of glutamines in the domain interface, which destabilizes the protein and leads to aggregation, while fluorescent quenching of tryptophans in the crystallin may have evolved to protect the protein from cataract-associated UV damage (Chen et al., 2006; Flaugh et al., 2006).

4. Amyloidosis

Domain swapping has been proposed as one mechanism by which amyloid fibers form. Amyloid fibers are insoluble protein aggregates that have been associated with a number of degenerative and deposition diseases. These diseases include Alzheimer's disease, Huntington's disease, Parkinson's disease, transmissible spongiform encephalopathies, senile systemic amyloidosis, type 2 diabetes, and hemodialysis-related amyloidosis (Chiti and Dobson, 2006). Although the proteins involved in each of these diseases have dissimilar amino acid sequences and native structures, they all involve the formation of unbranched fibrils that display green birefringence upon Congo red binding, thioflavin T binding, and a high degree of β-sheet content arranged in a cross-β structure that produces a characteristic X-ray diffraction pattern (Nilsson, 2004; Sunde and Blake, 1997).

The basic structure of amyloids is a cross-β structure, where β-strands are arranged perpendicularly to the fibril axis at a regular distance of approximately 4.8 Å from each other (Bonar et al., 1969; Eanes and Glenner, 1968; Sunde and Blake, 1997). X-ray crystallography of short, fiber forming peptides taken from larger amyloidgenic proteins (Nelson et al., 2005; Sawaya et al., 2007) and NMR of the Alzheimer's associated Aβ-(1-42) protein (Luhrs et al., 2005), have revealed atomic level details of the structure of amyloid fibers (Figure 1-1A). These structures and electron paramagnetic resonance spectroscopy have shown that a large number of amyloids form

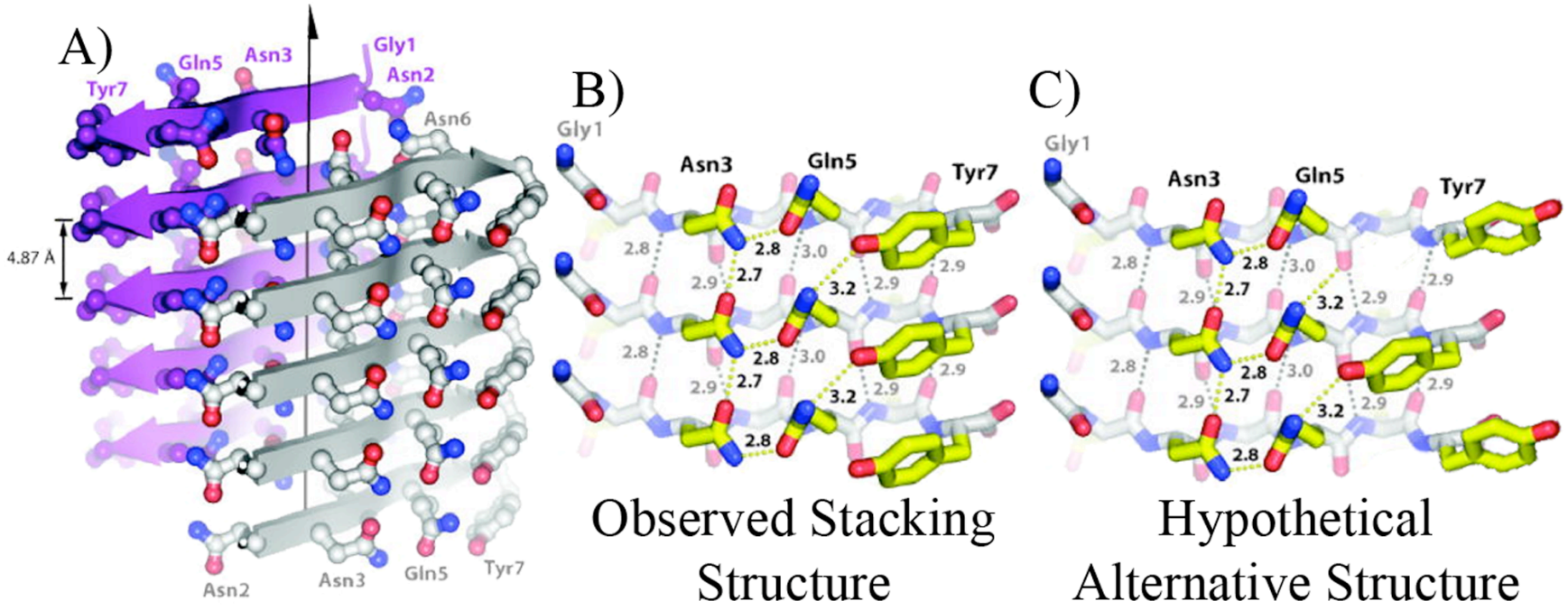

Figure 1-1. Side chain stacking in an amyloid crystal structure. (A) The crystal structure of an amyloidgenic seven-residue peptide from Sup35, demonstrating the typical cross-β structure. (B) Wet interface of the amyloid structure, demonstrating the polar and aromatic side chain stacks that dominates the side chain component of the structure. (C) A hypothetical alternative structure that demonstrates that the in-register parallel alignment of amyloids may be retained while side chain stacking in the observed aromatic stack could be disrupted by alternative side chain orientations. Adapted by permission from Macmillin Publishers Ltd: Nature (Nelson et al., 2005), © 2005.

in-register, parallel β-sheets, where identical residues from different polypeptide chains align in the β-sheets (Margittai and Langen, 2006). Alternatively, X-ray crystallography and isotope-edited IR spectroscopy demonstrate that some small peptide sequences form anti-parallel β-sheets, where specific alignments of β-strands in the β-sheets are observed (Petty et al., 2005; Sawaya et al., 2007; Silva et al., 2003).

A common feature of the parallel β-sheet amyloid structures is apparent in the positioning and orientation of side chains. Not only do identical residues align in-register with each other from one β-strand to the next in a β-sheet, but the side chains display identical orientations with respect to each other (Figure 1-1A&B). This is a phenomenon known as side-chain stacking, which will be described in more detail in Section C2. It is apparent from all of the current atomic resolution structures of parallel β-sheet amyloids that stacks, such as aromatic stacks, polar zippers, and hydrophobic stacks, dominate the side chain component of the amyloid structure (Luhrs et al., 2005; Nelson et al., 2005; Sawaya et al., 2007). It is important to note, however, that in-register, parallel alignment is a base criteria for side chain stacking, which further requires that side chains have similar chi angles or orientations with respect to each other (Petersen et al., 1997). A hypothetical example shows how alignment of tyrosines can be maintained while stacking is disrupted in a reimagined amyloid structure shown in Figure 1-1C.

Understanding the pathway that polypeptide chains follow in folding to these fibrillar structures is important for understanding the cause of cellular degeneration and disease, particularly because it has been proposed that the final fiber forms are not the toxic species in amyloidosis (Caughey and Lansbury, 2003). In this proposed model of toxicity, non-native oligomeric species that form “donut-like” shapes or early forming protofibrils may be toxic by either acting as membrane pores or by disrupting the function of membrane channels, while the formation of fibers may protect the cell against these toxic effects (Blanchard et al., 2004; Caughey and Lansbury, 2003). It is therefore important to investigate the kinetics and intermediates of the amyloidogenesis pathway.

Amyloid formation follows nucleation and growth kinetics according to aggregation experiments *in vitro* (Harper and Lansbury, 1997; Scheibel et al., 2004; Serio

et al., 2000). In the majority of nucleation-growth models for amyloid formation, the rate-limiting step for aggregation is the formation of an unstable nucleus. This nucleus may act as a template for further polymerization and conversion of other polypeptide chains (Petty and Decatur, 2005a; Scheibel et al., 2004). In aggregation kinetics, nucleation appears as a lag phase that precedes exponential growth of fibers. The exact mechanisms by which nucleation and conversion occur are highly debated.

Different amyloids appear to reach a nucleation state via different mechanisms. The soluble human prion protein is largely α-helical and must convert to the primarily β-sheet structure of the amyloid (Nguyen et al., 1995; Zhang et al., 1995). It appears that a refolding intermediate may be involved in this process, though the exact mechanism of conversion remains unclear (Prusiner, 1998). One model derived from a dimer structure of human prion protein suggests that domain swapping leads to the generation of an anti-parallel β-sheet that may be important for the transition to an amyloidgenic form (Knaus et al., 2001). Domain swapping has also been proposed as a mechanism for β-sheet formation in the cystatins, the T7 endonuclease I, and in the formation of amyloid fibers from an RNase A protein containing a poly-Q stretch (Guo and Eisenberg, 2006; Sambashivan et al., 2005; Sanders et al., 2004). Whether or not complete unfolding or partial unfolding is necessary for these domain swapping mechanisms is yet unclear.

In the case of transthyretin, amyloidogenesis proceeds from a partially folded monomer that is derived from the dissociation of the native tetramer (Colon and Kelly, 1992). Alternatively, a number of amyloids are derived from natively unfolded proteins, such as the NM region of the yeast prion Sup35 and the tau protein that is associated with amyloid deposits in Alzheimer's disease (Mukhopadhyay et al., 2007; Schweers et al., 1994). In these cases, the generation of nuclei either proceeds through oligomerization and conformational rearrangements, or from a rarely populated conformation that must proceed to oligomerization. FTIR studies on small peptides have revealed that β-sheet rearrangements and registration appear to be important steps in the nucleation and conversion of amyloidgenic peptides (Petty and Decatur, 2005a; Petty and Decatur, 2005b). This may be an important mechanism for nucleus formation independent of the folding state of the source of the amyloid.

Although most of the discussion above has concentrated on disease-related amyloidosis, not all amyloids are disease related. In yeast, prions have been shown to confer selective advantages under certain environmental conditions or stresses (True et al., 2004). Functional amyloids have been discovered in bacteria, where curli proteins form amyloid fibers to assist in biofilm formation (Chapman et al., 2002). Humans have functional amyloids, as best characterized in the case of Pmel17 whose amyloidogenesis is required for the polymerization of melanin (Fowler et al., 2006). Given the number of completely different proteins that are capable of forming amyloid under various environemental conditions, it has been proposed that the amyloid structure is a generic property of polypeptide chains that is independent of the amino acid sequence of the protein (Dobson, 1999).

Although the generic hydrogen bonding of the backbone may be the primary controller of amyloidogenesis, side chain identities play a role in the aggregation propensities of polypeptide chains (Kim and Hecht, 2005). Amino acid sequence also plays a role in the existence of species barriers in prion propagation and the specificity of cross-seeding reactions (Tessier and Lindquist, 2007). Examination of the relationship between amino acid sequence and folding pathways that lead to β-strand conformations both *in vitro* and *in vivo* may lead to curative or preventative agents for these diseases.

B. THE PROTEIN FOLDING PROBLEM

When Anfinsen originally demonstrated that RNase could refold when diluted from denaturants *in vitro*, he proposed that it could do so because the native state is the conformation with the globally lowest free energy for that particular amino acid sequence (Anfinsen, 1973; Anfinsen and Scheraga, 1975). According to the Levinthal paradox, if the polypeptide chain randomly sampled all possible conformations, then proteins could take the lifetime of the universe to reach their native state (Levinthal, 1968; Levinthal, 1969). In reality, proteins fold through folding pathways specified by their amino acid sequence that allows most proteins to fold on time scales from milliseconds to minutes both *in vivo* and *in vitro*. The essence of the protein folding problem is to understand how

a protein's sequence encodes folding pathways as well as the final structure.

Given the notion that the native conformation represent a globally minimal free energy, it has been proposed that many proteins may misfold and aggregate as discussed above due to the presence of kinitecally trapped local minima or alternative folding funnels in a protein sequences free energy landscape (Baker et al., 1992; Clark, 2004). The above discussion on diseases caused by misfolding and aggregation highlights the importance of understanding how the amino acid sequence encodes productive folding and avoids non-productive local free-energy minima. The discovery that severe neurodegenerative diseases like Alzheimer's, Huntington's, and prion diseases involve the generation of cross β-sheet structures demonstrates a particular importance for understanding the mechanisms of folding β-sheet proteins (Capaldi and Radford, 1998).

1. β-sheet Protein Folding

The sequence requirements for β-sheet formation are poorly understood. Recent Critical Assessment of Structure Prediction experiments (CASP5 and CASP6), have shown that current computational methods of predicting structure predict α-helices significantly better than β-strands when tested on sequences unrelated to known structures (Aloy et al., 2003; Vincent et al., 2005). The comparative success of α-helical predictions is in large part due to a number of biochemical studies that have elucidated sequence rules, such as the heptad repeat found in coiled coils (O'Shea et al., 1991; Woolfson and Alber, 1995), and general understanding for small, globular α-helical proteins like the four-helix bundles and globins (Kay and Baldwin, 1996; Kim and Baldwin, 1990). The elucidated sequence rules for α-helices have allowed for the design of new α-helical proteins (Ali et al., 2005; Lombardi et al., 1996; O'Shea et al., 1993). Without similarly detailed characterizations of β-sheet proteins, deciphering how amino acid sequences direct the folding of polypeptide chains will remain an unsolved problem.

One problem in elucidating the folding rules for β-sheet proteins lies in the fact that β-sheets are inherently more complex than α-helices. β-sheets are high contact order structures, meaning that hydrogen bonding or interacting residues are far apart in

sequence but are close in space in the final protein structure (Plaxco et al., 1998). In contrast, α-helices are low contact order structures where each residue in the helix hydrogen bonds with a residue only 4 positions away in the sequence. This low contact order allows for the chain to overcome the unfavorable entropic costs of nucleating the helix by having a high probability of formation due to the proximity of residues that must hydrogen bond (Clarke et al., 1999; Laurents and Baldwin, 1998). According to helix-coil transition theory, this nucleus is the most unstable transition state in the folding pathway. Once formed, the rest of the helix can rapidly zipper into the helical structure.

In contrast, the initial formation of β-structure may involve the interaction of residues that are hundreds of amino acids apart. This makes it more difficult to elucidate a generic pathway of folding for β-sheet proteins. Investigations on the folding of small β-sheet proteins, such as the Src homology 3 (SH3) domains of a number of proteins, have revealed that these β-sheet proteins appear to fold via a two-state mechanism (Capaldi and Radford, 1998). Phi analysis of small, fast β-sheet folders revealed that the formation of β-hairpins was primarily responsible for stabilizing the transition state, though this was not assumed to be sufficient evidence to state that hairpins initiated folding (Viguera et al., 1996).

In contrast to small β-sheet proteins, larger β-sheet proteins have been shown to populate partially folded intermediates. One example is the cellular retinoic acid binding protein I (CRABPI), which displays a β-clam topology where two β-sheets surround a central cavity (Thompson et al., 1995). Mutational studies on CRABPI, in conjunction with fluorescence spectroscopy and quench flow hydrogen exchange, have shown that the protein folds in four phases. These studies further showed that the substrate-binding pocket formed in the second phase while the protein's hydrogen bond network formed an order of magnitude slower in the third phase of folding, indicating that a near native confirmation could be obtained without the formation of stable hydrogen bonds (Clark et al., 1997; Clark et al., 1996). Gierasch and coworkers went on to demonstrate that β-turn and β-hairpin regions in CRABPI are encoding by local sequence rules, as two of the seven β-turns were able to form native-like structures when synthesized as peptides

(Rotondi and Gierasch, 2003a; Rotondi and Gierasch, 2003b). These β-turns participate in a network of conserved residues important for folding and stability and are likely amongst the first parts of the protein to fold (Rotondi et al., 2003). This indicates that the amino acid sequence of the β-turns conformationally restricts the structures of folding intermediates, driving the polypeptide chain towards its native structure.(Rotondi and Gierasch, 2003b; Rotondi et al., 2003).

Non-native-like folding intermediates have been observed as on-pathway folding intermediates for the β-trefoil protein interleukin-1β. Interleukin-1β rapidly collapses to a molten globule folding intermediate that possesses 90% of the native structure's β-sheet content, as discovered by circular dichroism (Ptitsyn et al., 1990). This collapse occurred on the order of milliseconds, while stable intermediates detectable by NMR did not appear until after 1 second of folding. Once stable intermediates were detectable with NMR, hydrogen-deuterium exchange showed that they had native-like β-sheets in two regions of the protein, indicating the formation of a more native-like folding intermediate (Varley et al., 1993). The authors suggested that these experiments indicate that the protein rapidly folds to a non-native β-sheet structure that slowly rearranges to a more native-like β-sheet structure. This indicates that β-sheets may fold through complex folding pathways that may include non-native-like conformations.

2. Protein Folding *In Vivo*

Due to the complexity of the cellular environment, protein folding studies have been almost strictly limited to *in vitro* refolding in the test tube. Since aggregation and off-pathway folding events have traditionally been considered nuisances rather than physiologically relevant, these *in vitro* experiments are often performed under conditions that minimize aggregation, such as low homogeneous protein concentrations and at low temperatures (Clark, 2004). These environmental conditions do not reflect physiological conditions in a living organism, where macromolecular crowding, heterogeneity, and cellular organization have profound effects on the folding and stability of proteins (Cheung et al., 2005; Minton, 2006).

Studies on the folding of the β-clam protein CRABPI have been extended into the context of the cell (Ignatova et al., 2007). Utilizing a fluorescence-based system known as FlAsH, the investigators determined the free energy of folding of the protein *in vivo*. The results demonstrated a close correlation between the *in vivo* and *in vitro* free energies of folding. However, kinetic rates for unfolding were altered in the cellular environment so that equilibrium was reached more quickly. The authors suggest that this may be due to the crowding effects of the cell.

Fluorescence-based methods have further demonstrated that the ribosome interacts with the nascent chain it is translating to induce conformational changes (Woolhead et al., 2006). In turn, the nascent chain encodes sequences that interact with the ribosome to control the rate of translation.

These experiments highlight the fact that co-translational folding of nascent chains occurs in a complex and interactive environment that is not accurately reproduced by refolding experiments out of denaturant. To truly uncover how the sequence of a polypeptide chain encodes the folding and structure of a protein in its *in vivo* environment, complementary *in vivo* and *in vitro* experiments must be performed on model protein systems. This is of particular importance for β-sheet proteins, which appear to have complex folding pathways that are often aggregation prone.

C. THE β-HELIX FOLD

The parallel β-helix, a structurally repetitive fold created by the coiling of the polypeptide backbone, is one of the topologically simplest β-sheet folds (Jenkins and Pickersgill, 2001; Jurnak et al., 1994; Pickersgill et al., 1998; Yoder et al., 1993a). Each coil or rung of the canonical parallel β-helix fold consists of three β-strands (A, B and C) interrupted by variable turn or loop regions (T1, T2, and T3) (Figure 1-2A) (Jurnak et al., 1994). In most of the parallel β-helix proteins, loops may be large in sequence and may contain entire folding domains. Rungs are aligned, forming a cross-β structure of elongated β-sheets lying parallel to the helical axis. Structural repetition of coils creates a

cylindrical hydrophobic core. β-helices are often capped by α-helix motifs or anti-parallel β-strands in order to avoid exposure of the hydrophobic core and unfulfilled hydrogen bond donors and acceptors at the ends of the helices (Richardson and Richardson, 2002)

Three main subclasses of parallel β-helix fold currently exist: the right-handed parallel β-helix, the left-handed parallel β-helix, and the triple-stranded parallel β-helix. The right-handed β-helix proteins coil around the axis of the helix in a right-handed fashion and are further characterized by the presence of a kidney bean shaped cross-section (Figure 1-2B) (Jenkins and Pickersgill, 2001). In contrast, both the left-handed and triple-stranded β-helices wrap around the helical axis in a left-handed direction and have a cross-section that is triangular in cross-section (Figure 1-2C) (Jenkins and Pickersgill, 2001; Weigele et al., 2003). Examples of these distinct classes of β-helix folds are shown in Figure 1-3.

A number of monomeric and trimeric parallel β-helix proteins display an unusual stability, whereby the proteins have extremely high melting temperatures, are resistant to protease digestion, and are stable in the presence of detergents (Jenkins and Pickersgill, 2001; Weigele et al., 2003)

As of this writing, there are 168 β-helix structure files deposited amongst the 44,018 known structures in the protein data bank (PDB) (Berman et al., 2000). These β-helix structures represent approximately 40 different protein classes. Overall, the β-helix fold is not a common structural fold (Bradley et al., 2001). The sequences of these proteins are quite diverse, displaying typical sequence similarities between any two β-helix proteins of less than 15% (McDonnell et al., 2006).

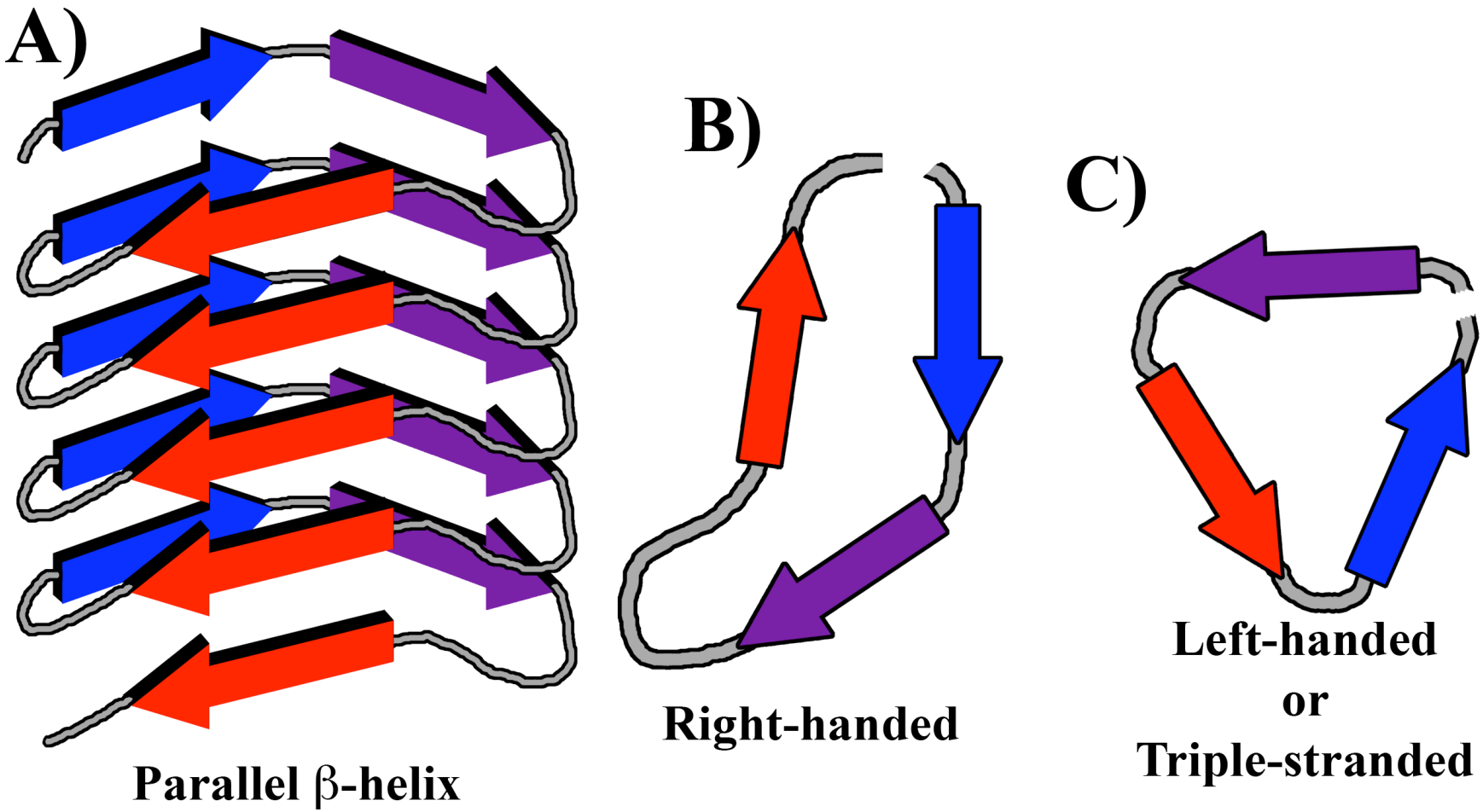

Figure 1-2. Topology of the parallel β-helix. (A) Ribbon diagram of the canonical β-helix topology, which consists of three β-sheets, which are each colored individually. (B) Flattened ribbon representation of the kidney bean shaped cross-section of a right-handed β-helix. (C) Flattened ribbon representation of the triangular cross-section of a left-handed β-helix or a triple-stranded β-helix.

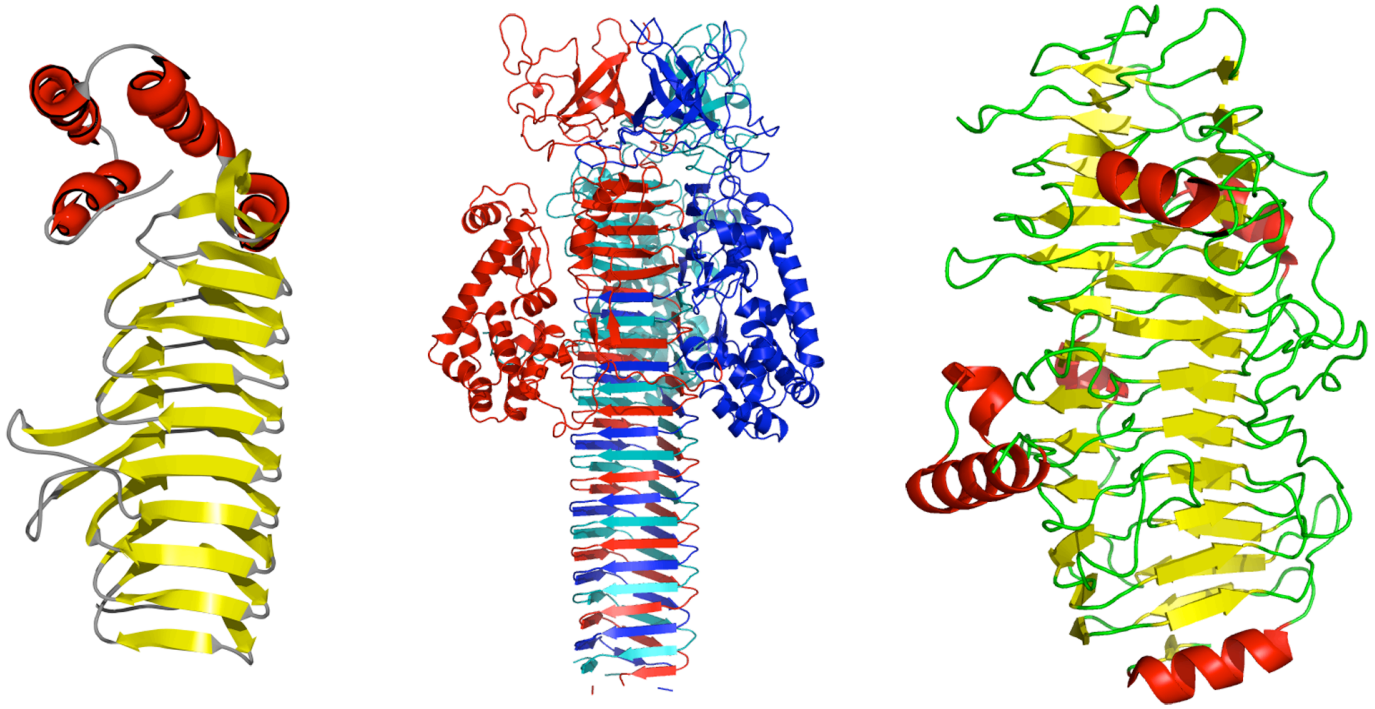

Figure 1-3. Sample parallel β-helix protein structures. Ribbon diagrams of three β-helix proteins. The left-handed UDP N-acetylglucosamine acyltransferase LpxA protein (PDB ID: 1LXA) (Raetz and Roderick, 1995) and the right-handed Chondroitinase B (PDB ID: 1DBG) (Huang et al., 1999) are monomeric proteins that are colored according to their secondary structure. The gp5 component of the T4 lysozyme complex (PDB ID: 1K28) (Kanamaru et al., 2002) is a homotrimer where each polypeptide chain is individually colored red, blue, or cyan to highlight the interdigitation of the three chains.

1. β-helix Functions

Known β-helical proteins are predominantly associated with disease or infection. Examples include a number of viral adhesin proteins, virulence factors such as the pertactin and hemagglutinin of *Bordetella pertussis*, and the major pollen allergen from the mountain cedar tree (Clantin et al., 2004; Czerwinski et al., 2005; Emsley et al., 1996; Jenkins and Pickersgill, 2001). A non-canonical four-sided β-helix fold, which displays a square cross-section, has recently been discovered that confers antibiotic resistance to *Mycobacterium tuberculosis* by mimicking the structure of DNA in order to protect DNA gyrase against fluoroquinolone inhibition (Hegde et al., 2005). Human retinitis pigmentosa 2 protein (RP2), mutations of which are the cause of 10 – 15% of X-linked retinitis pigmentosa (RP) cases (Hardcastle et al., 1999; Schwahn et al., 1998), contains a right-handed parallel β-helix domain. A large number of mutations associated with RP are located in the β-helix domain, many of which are in the buried core of the β-helix (Kuhnel et al., 2006).

The most common functional theme in the β-helix proteins is the use of an elongated lateral surface to recognize an extended polysaccharide sequence. (Jenkins and Pickersgill, 2001). The structures of P22 tailspike, pectate lyase C, and chondroitinase B have all been solved in the presence of polysaccharide ligand and demonstrate the use of a cleft on one of the β-sheet faces of the β-helix to bind ligand, though the angle of this binding with respect to the axis of the helix depends on the particular protein (Huang et al., 1999; Scavetta et al., 1999; Steinbacher et al., 1997b). An alternative mode for polysaccharide binding has been observed in the cleft between monomer subunits of the trimeric β-helix inulin fructotransferase of *Bacillus sp. snu-7* (Jung et al., 2007). In the crystal structures of ι-carragenase with and without a cell wall polysaccharide ligand, the β-helix surface acts as part of a binding cleft but the extended loops that emerge from the turn regions of the β-helix fold over the substrate to form a catalytic tunnel (Michel et al., 2003). This motif of binding long molecules is not limited to polysaccharides. The non-canonical two-sheet left-handed β-helical collagen binding domain of YadA from *Yersinia enterocolitica* appears to bind human collagen in a similar manner along the long lateral surface of its β-helix (Nummelin et al., 2004).

YadA from *Yersinia enterocolitica*, P.69 pertactin from *Bordetella pertussis*, and hemoglobin protease from *E. coli* are all β-helical autotransporters, whose folding and toxicity is coupled to membrane translocation (Emsley et al., 1996; Otto et al., 2005; Roggenkamp et al., 2003). The β-helical filamentous hemagglutinin from *B. pertussis* must also couple its folding to membrane translocation, but does so through a two-partner secretion pathway (Clantin et al., 2004). In both transport pathways, outer membrane translocation is ATP-independent and proton-gradient independent (Henderson et al., 2004; Thanassi et al., 2005). It has been proposed that in the case of pertactin, a vectorial folding of the β-helix structure may contribute the requisite energy for efficient energy-independent translocation (Figure 1-4) (Junker et al., 2006). Greater than 97% of autotransporters have been predicted to be β-helices, implicating that the energetic coupling between folding and secretion may be a more general phenomenon (Junker et al., 2006).

Multiple computational methods have predicted the existence of hundreds of additional β-helical proteins that are largely associated with infectious disease, pathogens, or allergens (Bradley et al., 2001; Ciccarelli et al., 2002; Jurnak et al., 1994; McDonnell et al., 2006). Given the rarity of currently known and predicted human β-helical proteins, this fold represents a potential target for novel therapies against pathogens. Since a number of these potential targets are likely to fall into the family of autotransporting virulence factors, not only is the native state a target for therapeutics but so are the folding intermediates that contribute to the virulence factor's secretion from the pathogenic cell. It is therefore critical to not only know the native state of the β-helix proteins, but to understand their sequence-dependent folding pathways.

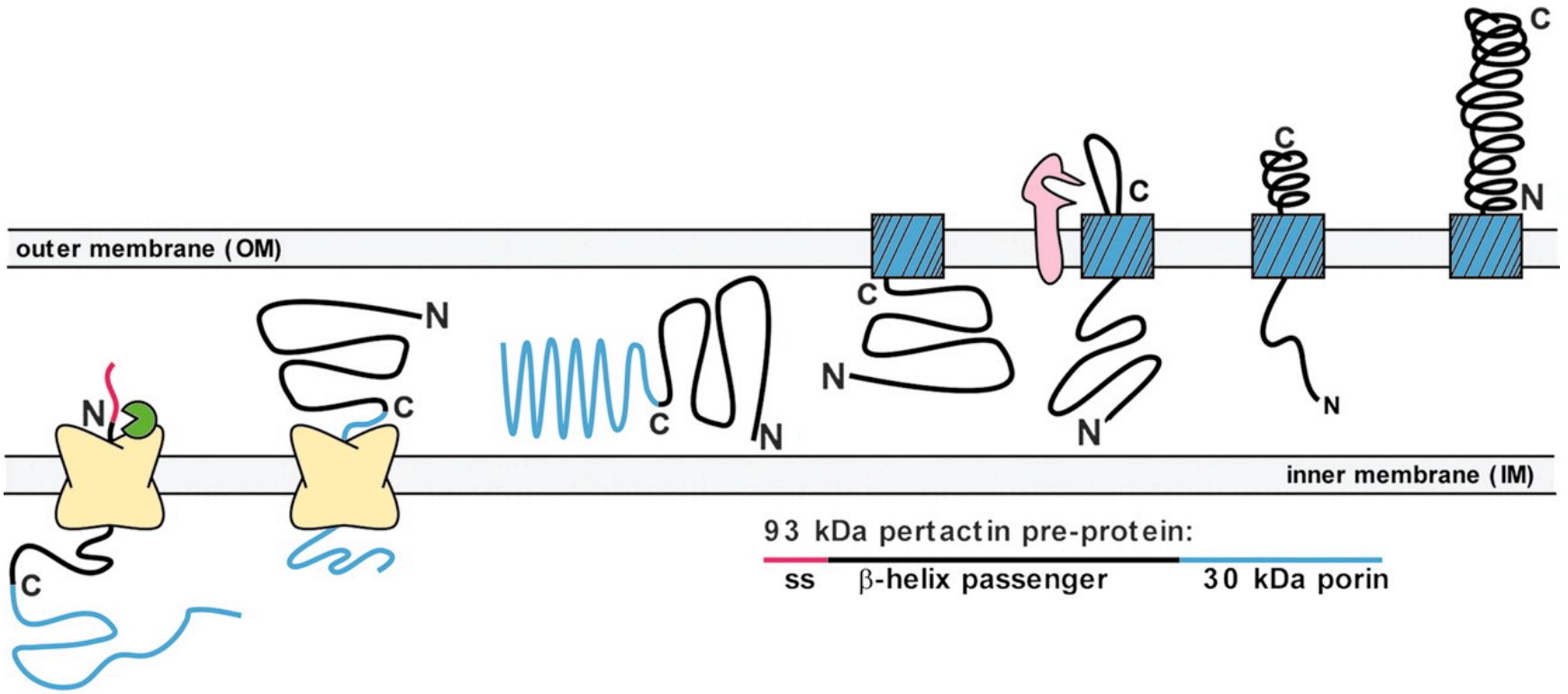

Figure 1-4. Model for the autotransport and folding of pertactin. In the model proposed by Junker, et al., pertactin is secreted into the periplasmic space via an N-terminal signal sequence (pink). Once passaged across the inner membrane, the 30 kDa C-terminal porin sequence (blue) forms in the outer membrane. The passenger domain subsequently passes through the porin complex and folds into its native β-helical structure on the outside of the cell. Figure reprinted from (Junker et al., 2006), © 2006 by The National Academy of Sciences of the USA.

2. Side Chain Stacking

When the β-helix fold was first discovered upon solving the crystal structure of pectate lyase, it was noted that the side chains of the β-helix structure formed linear stacks that included asparagines ladders, serine stacks, aliphatic stacks, and aromatic stacks (Yoder et al., 1993a; Yoder et al., 1993b). As more β-helix protein structures were solved, it became apparent that the existence of elongated stacks composed of aligned, similar side chains with similar orientations was a prominent feature common to all β-helical proteins (Jenkins and Pickersgill, 2001). "Cupped stacks" of aliphatic residues, aromatic stacks, and polar zippers that are nearly identical in appearance to the side chain stacks observed in amyloid structures (Figure 1-5C, Figure 1-1, Section A4) have all been observed in β-helical structures (Jenkins and Pickersgill, 2001; Petersen et al., 1997; Raetz and Roderick, 1995). While side chain stacks in the currently solved amyloid structures are always composed of identical residues, the side chain stacks observed in β-helices are often composed of similar side chains that continue to exhibit similar orientations, in spite of a lack of sequence identity. Although stacks are composed of similar residues, structurally repeated rungs do not exhibit an overt sequence repeat nor do they exhibit sequence similarity across β-helical proteins (Bradley et al., 2001). This is particularly true for the right-handed parallel β-helix proteins, while the left-handed β-helix proteins sometimes display a mild sequence repeat. In the right-handed β-helices, identical residue stacks of more than four rungs in length are rarely observed.

It should be noted that stacking of residues in parallel, particularly aromatic residues, is not a common feature in the buried cores of globular proteins. In the case of aromatics, edge-to-face orientations are more common in non-β-helix proteins (Burley and Petsko, 1988). In the context of the β-helices, aromatics do stack as if mimicking the stacking of bases in DNA. This can be seen in an aromatic stack composed of twelve aromatics in the buried core of the chondroitinase B protein (Figure 1-5C, left).

When examining the structures of β-helix proteins in the protein database, I noticed that side chains that point into the core of the β-helix tend to be aligned in well-organized stacks, where as those side chains that point out into the solvent tend to align

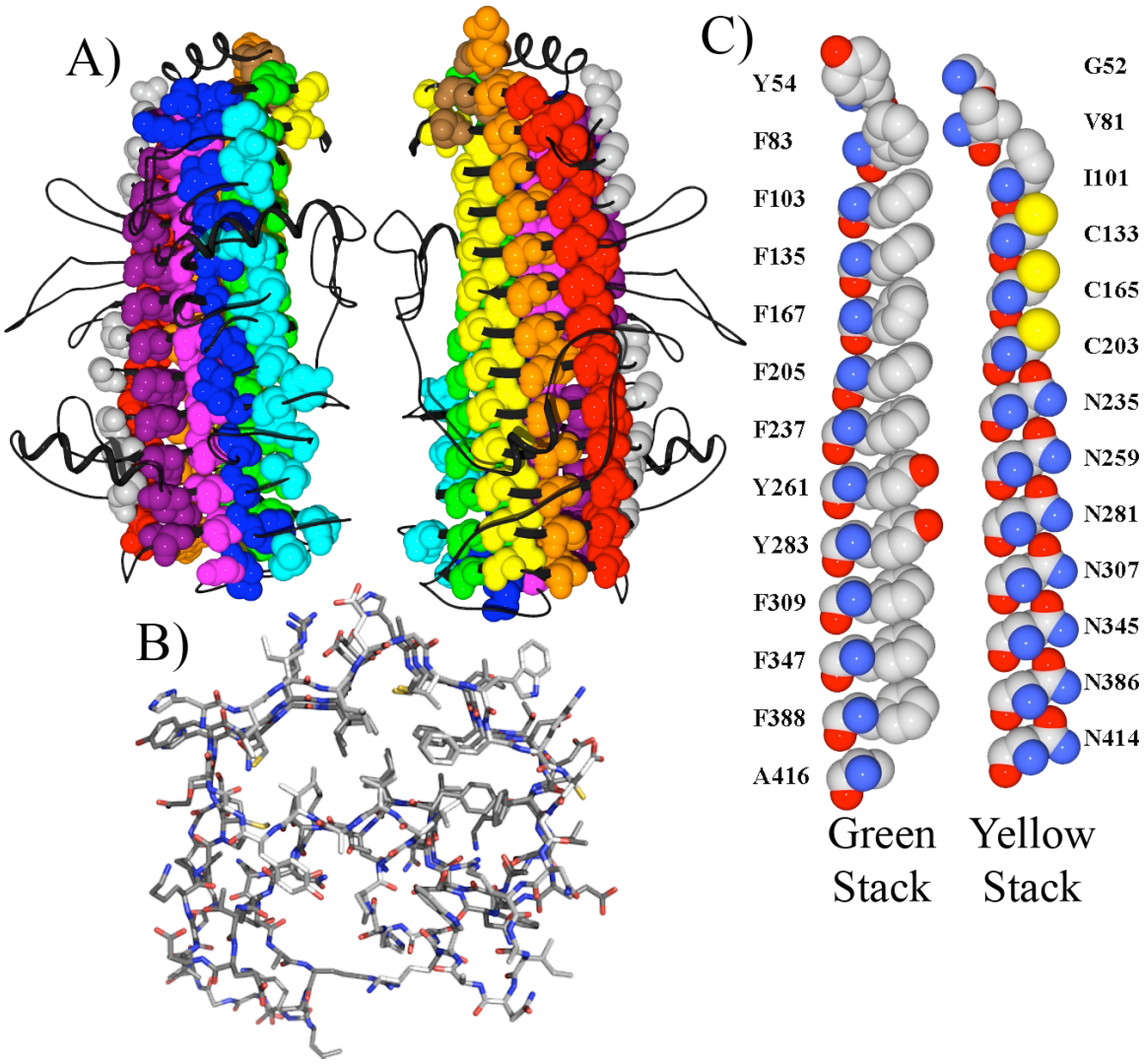

Figure 1-5. Side chain stacking in the buried core of chondroitinase B. (A) Inward-pointing, buried core residues are shown in sphere representation on top of the ribbon diagram of the chondroitinase structure (PDB ID: 1DBG) (Huang et al., 1999). Each stack is colored individually. The image on the right is rotated by 180º around the helix axis relative to the left image. (B) Cross-section of rungs 3 – 5, demonstrating the organized stacking of side chains in the core, as opposed to the solvent exposed exterior. (C) The green and yellow buried core stacks in (A) are shown in sphere representation absent the rest of the structure.

poorly in their side chain orientations (Figure 1-5B). We found this to be particularly true for the right-handed β-helices. The organization of these buried core residues is likely an indicator of their packing constraints, which may indicate that they play an important role in encoding the conformation of the protein.

One notable exception to this observation is in the group of insect anti-freeze proteins, which are left-handed β-helices. These proteins display a high degree of well-ordered side-chain stacks on the exterior of the protein. In particular, one β-sheet of the spruce budworm anti-freeze protein displays nearly perfect threonine stacks on its exterior face (Graether et al., 2000). It has been proposed that this stacking has an evolved function, whereby the geometry of the threonine stacks mimic the lattice of water molecules in ice, thereby allowing the protein to bind ice crystals and prevent further crystallization.

Side chain stacking is a dominant feature of both β-helices and amyloids. It is unclear how this apparent one-dimensional crystallization contributes to the folding of these proteins, but it is clear that it plays a crucial role in their final structure. Understanding the contributions of side chain stacking in the context of the β-helices may lend clues to its role in amyloids. Due to a number of common features, the β-helix fold has been proposed as a model for that of full-length amyloid protein fibers (Kishimoto et al., 2004; Lazo and Downing, 1998; Perutz et al., 2002; Wetzel, 2002; Williams et al., 2004). In one such model, the NM region of the yeast prion Sup35 has been proposed to form β-helical units that interact in a tail-to-tail and head-to-head fashion in order to form a fiber (Figure 1-6A) (Krishnan and Lindquist, 2005). A head-to-head assembly has been observed in the crystal structure of the globular, soluble potato ADP-glucose pyrophosphorylase (Figure 1-6B&C) (Jin et al., 2005), indicating that experimentation into the folding and structure of the β-helix proteins may lead to a better understanding of β-sheet formation and the assembly of higher order structures that underlies amyloidogenesis.

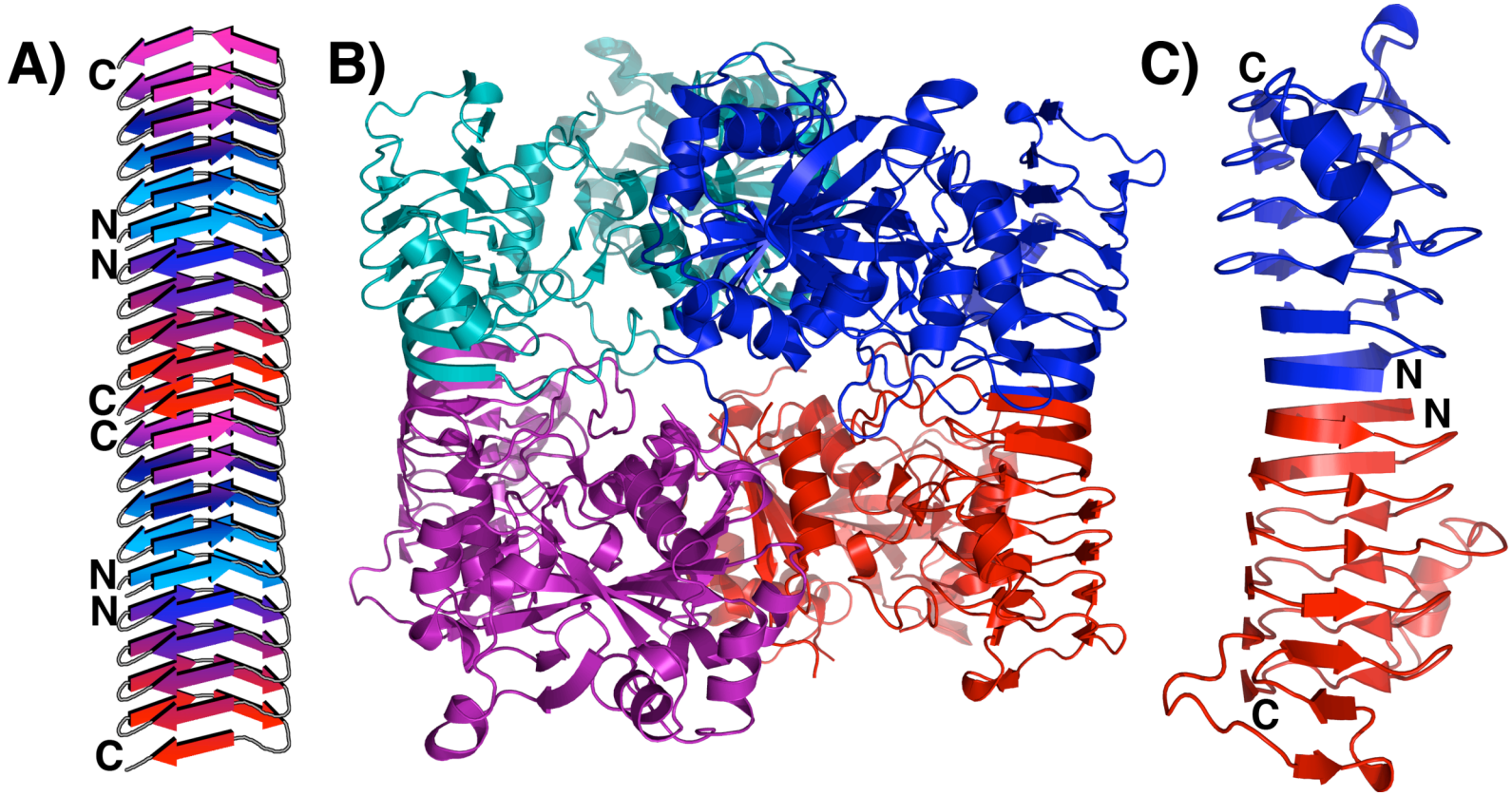

Figure 1-6. Head-to-head stacking in β-helices is a model for amyloid. (A) A diagram of the β-helical model for the amyloid fiber generated by the NM region of Sup35 proposed by Krishnan and Lindquist (Krishnan and Lindquist, 2005). β-helical units are colored in a gradient from their N-termini (blue or cyan) to their C-termini (red or magenta) to highlight the anti-parallel head-to-head and tail-to-tail interactions. (B) Ribbon diagram of the tetrameric potato tuber ADP-glucose pyrophosphorylase (PDB code: 1YP2), where each polypeptide chain is colored individually (Jin et al., 2005). (C) The β-helical domains of the red and blue chains in (B) are shown without the rest of the structure to show the head-to-head stacking of the β-helices that form part of the intermolecular interface in the protein's tertiary structure. N = N-terminus; C=C-terminus.

D. P22 TAILSPIKE

1. Structure of the P22 Tailspike

The tailspike protein of *Salmonella* phage P22 has been one of the principal systems for studying protein folding in β-helices (Jenkins and Pickersgill, 2001). The 2 kilobase long *gene 9* in the phage P22 genome encodes the tailspike's 72kDa polypeptide chain (Sauer et al., 1982). Three of these 666-residue polypeptide chains come together to form the multi-domain native structure of the tailspike (Figure 1-7A) (Steinbacher et al., 1996; Steinbacher et al., 1997a; Steinbacher et al., 1994). At the N-terminus of the homotrimer is the head-binding domain, which allows the tailspike protein to assemble onto the tail machine of the phage particle but is not necessary for the folding of the rest of the protein (Chang et al., 2006; Danner et al., 1993; Lander et al., 2006).

Subsequent to the N-terminal head-binding domain, each polypeptide chain contains a 13 rung, right-handed parallel β-helix domain (residues 143-540). This β-helix domain includes the globular dorsal fin domain, which sticks out of the β-helix at turn T3 of rung 3 (Figure 1-7B). Tailspike's β-helix is capable of recognizing the O-antigen of *Salmonella's* cell surface lipopolysaccharide and cleaving it via an endorhamnosidase activity (Seckler, 1998). Upon binding, the long, floppy O-antigen sits in the elongated lateral cleft of the β-helix, located between the dorsal fin and the ventral fin (Figure 1-7C). This binding mode brings the polysaccharide into close contact with the active site residues D392, D395 and E359, which are capable of cleaving the α(1,3)-O-glycosidic bond between rhamnose and galactose in the O-antigen (Steinbacher et al., 1997b). Although this endorhamnosidase activity is not necessary for phage adsorption, it is believed to help bring the phage to the cell membrane so that DNA injection can occur (Berget and Poteete, 1980; Israel, 1978; Iwashita and Kanegasaki, 1976).

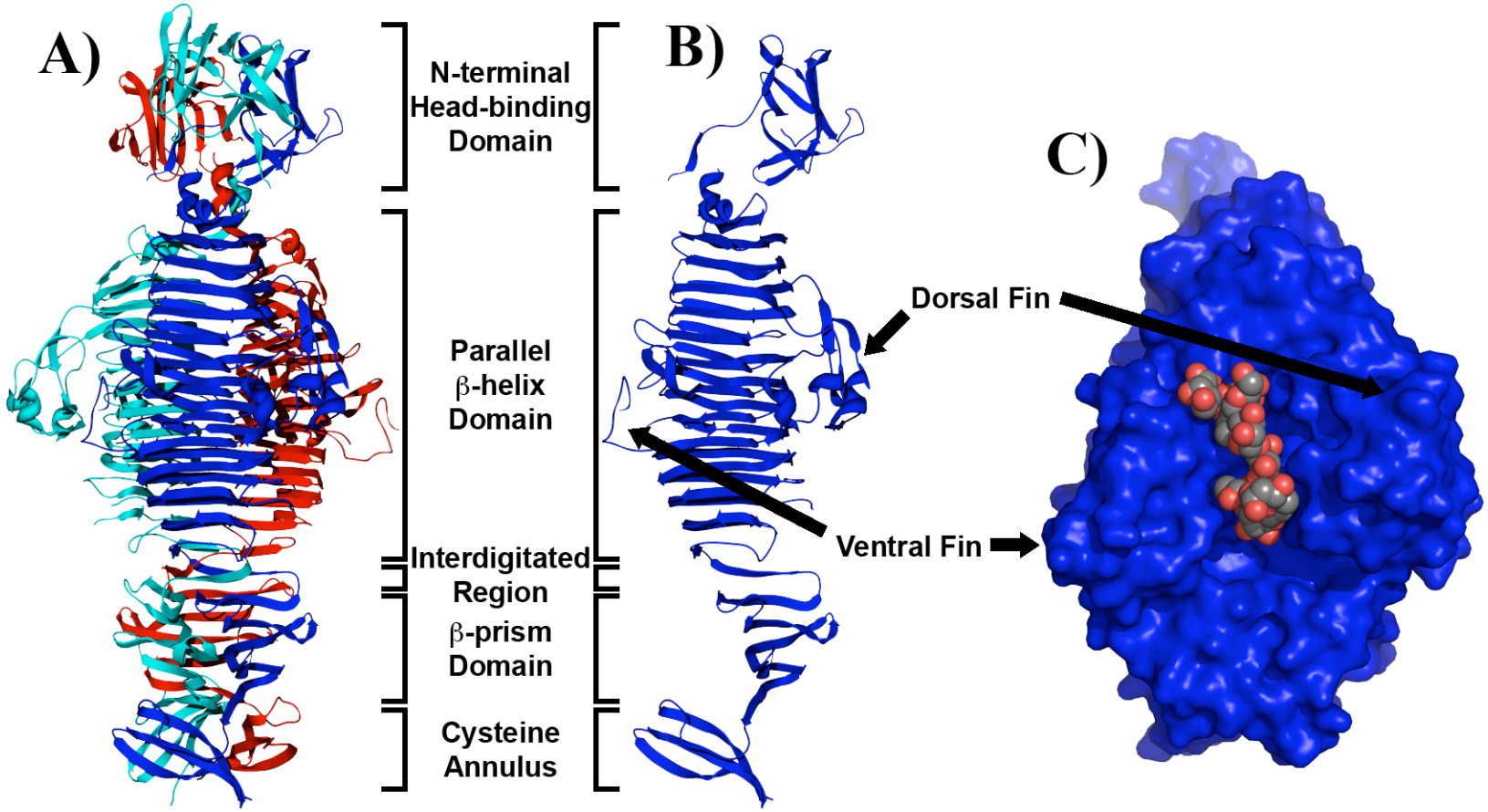

Figure 1-7. Structure of the P22 tailspike protein. (A) Ribbon diagram of the P22 tailspike multi-domain homotrimer. The N-terminal domain (PDB code 1LKT) is shown above the C-terminal part of tailspike (PDB code 1TYU), though the two parts were crystallized separately (Steinbacher et al., 1996; Steinbacher et al., 1997a). (B) Ribbon diagram of a single polypeptide chain pulled out of the native trimer structure. The domains, the dorsal fin, and the ventral fin are labeled. (C) The molecular surface of the β-helix domain is shown binding a fragment of *Salmonella*'s O-antigen polysaccharide (VDW representation) in the binding cleft between the dorsal fin and the ventral fin.

C-terminal to the β-helix domain, monomers wrap around each other to form an intertwined β-sheet domain where each chain's residues contribute to a single buried hydrophobic core, consisting of one rung of a triple β-helix (residues 541-557) and a triple β-prism structure (Kreisberg et al., 2000; Seckler, 1998; Steinbacher et al., 1994). The ultimate step in tailspike's productive folding pathway (Figure 1-8) coincides with the folding and assembly of this interdigitated region to form a molecular clamp that makes tailspike resistant to trypsin degradation and SDS denaturation, and raises the melting temperature by approximately 40°C to 88.4°C (Goldenberg and King, 1982; Kreisberg et al., 2002; Sturtevant et al., 1989). Because of the native structure's SDS resistance, final native structure can be easily separated and identified from folding intermediates and aggregates by SDS gel electrophoresis.

At the far C-terminus of the homotrimer is a region of the protein called the cysteine annulus (Figure 1-7B), so called due to the total presence of six cysteine residues, two (C613 and C635) from each polypeptide chain.

2. *In Vivo* Folding Pathway of P22 Tailspike

Extensive studies of the *in vivo* folding pathway of tailspike have revealed a number of long-lived, partially folded intermediates (Figure 1-8) (Betts and King, 1999). The first of these intermediate structures, a β-helical conformer [I_0], rapidly develops while nascent chains are still attached to the ribosome (Clark and King, 2001). Upon release from the ribosome tailspike, thermolabile monomeric intermediates, [I], may oligomerize and further fold into the SDS-resistant trimeric tailspike, NT (Goldenberg et al., 1982).

Oligomerization of the tailspike into a natively structured heterotrimer appears to involve the region of the protein C-terminal to the β-helix (Gage and Robinson, 2003). Part of this assembly reaction involves the formation of a transient disulfide bond that is not present in the final native structure (Danek and Robinson, 2003; Haase-Pettingell et al., 2001; Robinson and King, 1997). Initial evidence suggests that this disulfide bond may first form in the monomeric folding intermediate [I] (Kim and Robinson, 2006),

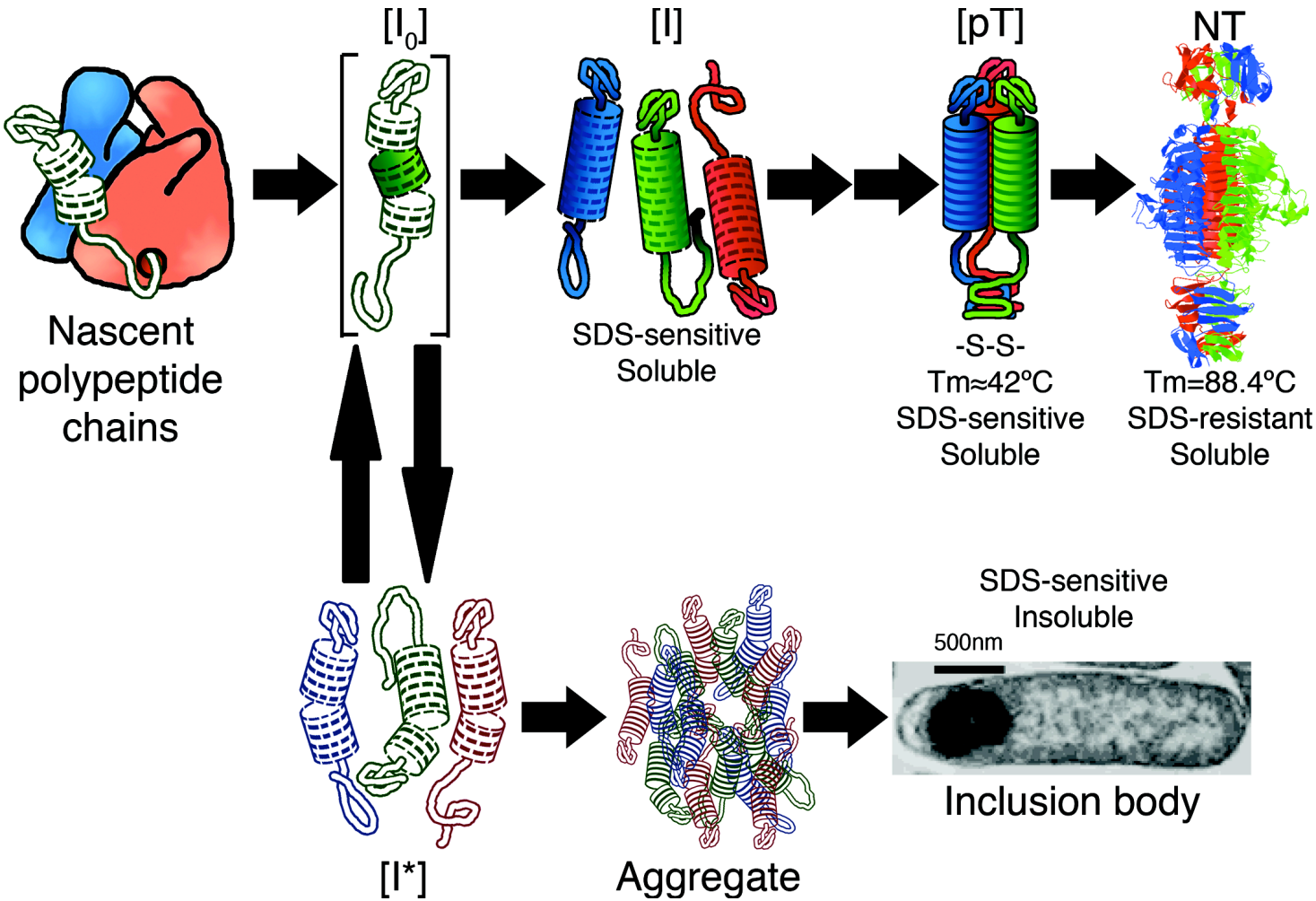

Figure 1-8. *In vivo* folding pathway of the P22 tailspike homotrimer. As described in the text, the upper branch of the folding pathway shows productive intracellular intermediates leading to the native state. Intermediates leading to the inclusion body state are on shown on the bottom. [pT] = Protrimer.

where it may aid in the formation of an intermolecular disulfide bond in the dimer and protrimer intermediates (Betts and King, 1998; Robinson and King, 1997). The transition from the protrimer to the native trimer involves the reduction of transient disulfide bonds and is coincident with further folding of the interdigitated and β-prism regions of the protein (Kreisberg et al., 2002). Mutations have been isolated in these regions that block or delay this final transition to the native trimer (Gage et al., 2005; Weigele et al., 2005). From such mutants, it has been shown that mutations that block oligomerization or trimer maturation result in the accumulation of soluble, SDS-sensitive species (Betts et al., 2004; Kreisberg et al., 2002).

Alternatively, the presence of any of over 60 genetically isolated temperature-sensitive folding (*tsf*) mutations at temperatures in the higher range of cellular growth will generate a misfolded monomeric conformation [I*], which specifically self-aggregates and accumulates in inclusion bodies (Haase-Pettingell and King, 1997; Haase-Pettingell and King, 1988). Higher temperatures prevent the wild-type tailspike protein from folding successfully, causing wild-type chains to aggregate in the cell. This in turn limits the growth of the P22 phage (Pope et al., 2004). Aggregation proceeds via monomer addition of aggregation prone intermediates that are conformationally similar to productive folding intermediates (Speed et al., 1997; Speed et al., 1995). Though a specific aggregation pathway exists, the conformation of polypeptide chains in these aggregates is unknown and may contain native-like conformations.

The *tsf* mutations, predominately located at surface loops and turns, appear to act by destabilizing the β-helical structure found in the thermolabile, monomeric folding intermediate (Haase-Pettingell and King, 1997; Schuler and Seckler, 1998). Global suppressors (*su*) of the *tsf* mutations or lower temperatures stabilize the β-helical fold or otherwise inhibit aggregation (Danner and Seckler, 1993; Schuler and Seckler, 1998). The GroEL chaperone does not assist tailspike chains in efficiently folding or preventing aggregation *in vivo* (Gordon et al., 1994). Partitioning between native and aggregation states is therefore a reflection of the folding state of the β-helix. As a result of the native tailspike's SDS-resistance, SDS-PAGE of whole lysates provides a sensitive and reliable measure of *in vivo* folding (Goldenberg and King, 1982; Goldenberg et al., 1982).

3. Stacking in the P22 Tailspike

The numerous surface site *tsf* mutations represent residues essential for the creation of the β-helical fold at higher temperatures but do not carry the essential information for β-helix folding, as they are unnecessary at lower temperatures. In contrast, five phenylalanines and one leucine in the hydrophobic core stacks are essential for the folding, but not stability, of tailspike at high and low temperatures (Betts et al., 2004). This suggests that stacked core residues contain the necessary sequence information to direct β-helix formation.

It has been observed that disruptions of hydrophobic side chain packing in the cores of β-sheet and β-barrel proteins disrupts their folding or destabilizes the native state (Gunasekaran et al., 2001; Sauer, 1996). This implies that the buried core residues of the P22 tailspike may be particularly important in encoding the β-helix structure. We have observed that the buried core of the P22 tailspike is dominated by the presence of side chain stacks (Figure 1-9). In total, we identified ten main columns of stacks, each of which is displayed in its own individual color in Figure 1-9. These stacks encompass 113 buried core residues, six of which are the amino acids discovered to be essential for folding *in vivo* (Betts et al., 2004). Like the observations made on other β-helix proteins, these stacks appear better organized in relation to shared side chain orientations than residues that point out into the solvent. These observations further indicate that these buried core, side chain stacking residues are primary candidates for the residues that encode the folding and structure of the P22 tailspike protein.

Investigating the role that buried, side chain stacking residues play in the folding of the P22 tailspike's β-helix domain may yield information relevant to amyloid formation, especially since a fragment composed solely of the isolated β-helix domain is capable of forming amyloid fibers (Schuler et al., 1999).

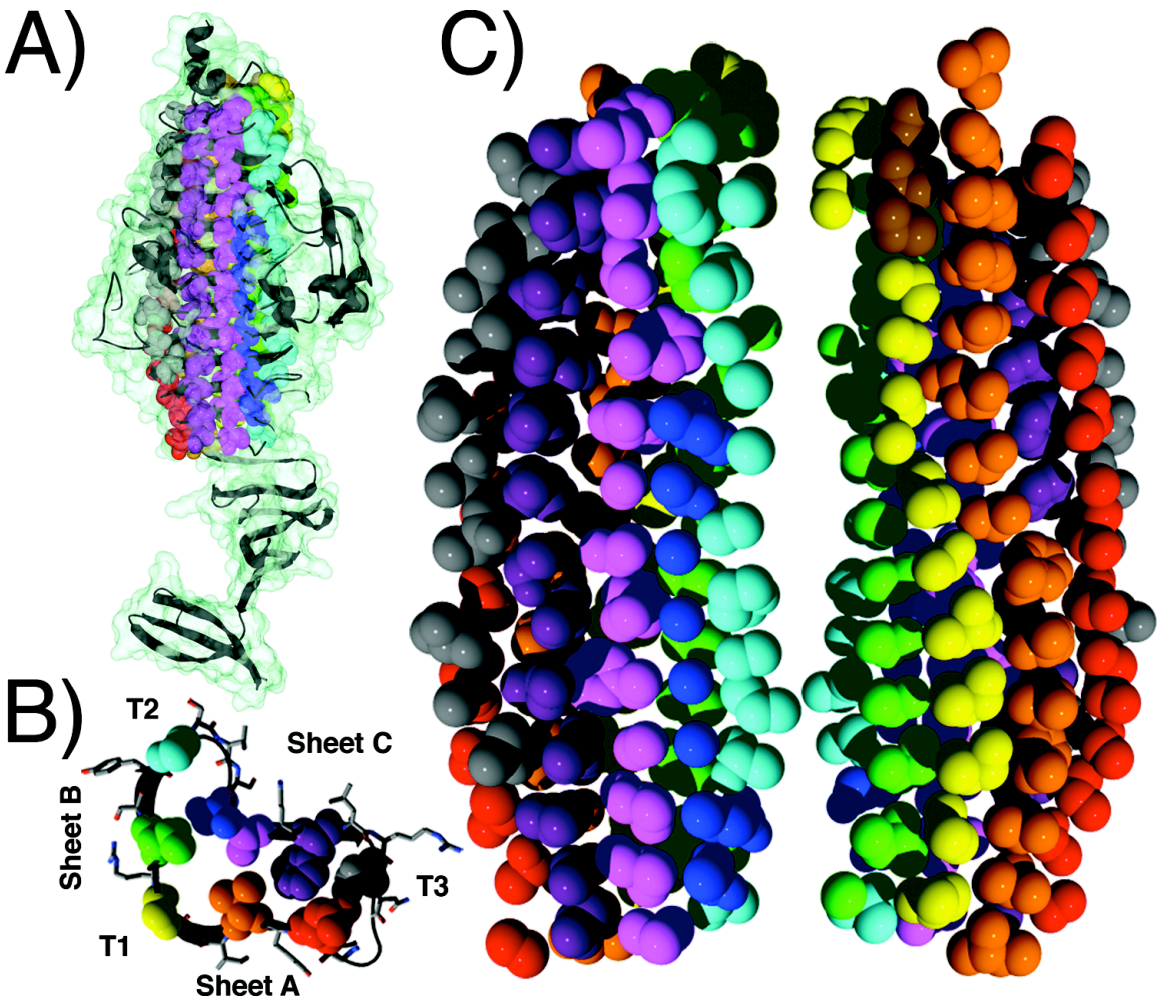

Figure 1-9. Side chain stacking in the core of the P22 tailspike β-helix. (A) The C-terminal portion (113-666) of a single chain of the P22 tailspike trimer is shown as a ribbon diagram and a transparent molecular surface. Backbone and side chain atoms of inward pointing, buried core residues are shown in space-fill and colored according to their stack. Solvent exposed residues are not shown for clarity. (B) Rung 6 cross-section of the β-helix domain, demonstrating the locations of inward pointing stacks (spheres) and outward pointing residues (sticks). (C) Side chain atoms of the buried core. Single spheres represent glycines. The left image is in the same orientation as (A); the right is rotated by 180º. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

E. THESIS OBJECTIVES

The aim of this thesis is to elucidate the role that the amino acid sequence of a protein plays in the importance of the side-chain stacking phenomenon. Specifically, this thesis investigates the contributions of the amino acid sequence of the buried core stacks of the β-helix domain of P22 tailspike towards the folding of the β-helix. The roles of these buried side chains was first investigated through a large-scale scanning alanine mutagenesis experiment (Chapter 2). The results of this mutagenesis experiment lead to a model pathway for the folding of the β-helix domain, which is subsequently tested and examined in the remaining chapters.

These studies lead to a better understanding of the tailspike's folding pathway, with particular attention to the folding of the β-helix domain. In the broader scheme, the results of the thesis contributes understanding into the mechanisms by which an amino acid sequence may direct polypeptide chains into a folded state or into aggregation. Given the similarities between tailspike and other aggregation prone proteins, such as amyloids, this thesis has general application to our understanding of β-sheet folding, assembly, and aggregation.

CHAPTER TWO:

IN VIVO MUTAGENESIS OF THE P22 TAILSPIKE β-HELIX BURIED CORE

A. INTRODUCTION

1. Stacking in the Buried Core of the P22 Tailspike

Upon inspection of the structure of the tailspike protein (PDB file 1TYU (Steinbacher et al., 1996)), it became apparent that the majority of the β-helix buried core is composed of stacked residues whose side chains have similar conformations (Figure 1-9). In total, ten buried core stacks were identified containing 113 amino acids that comprised the entirety of the β-helix domain's core: six stacks participating in the structurally conserved β-sheets (red, orange, yellow, green, magenta, and purple stacks in Figure 1-9), two pseudo-stacks located in T2 and T3 (cyan and gray), one short stack preceding sheet C running from rungs 5 to 13 (blue), and a second shortened stack of three residues at the N-terminus of the β-helix in the conserved turn T2 (brown).

2. Experimental Design

Side chain stacking is a prominent feature of the buried core of the P22 tailspike β-helix domain, in spite of the fact that this domain does not display an overt sequence repeat (Figure 2-1). How does the amino acid sequence of these stacks contribute to the folding of the protein and how are residue mismatches in stacks tolerated?

To investigate the contributions of all stacked, buried core side chains to the folding of the parallel β-helix, I have carried out a high-throughput, scanning alanine mutagenesis study on the folding of the P22 tailspike (Cunningham and Wells, 1989). Scanning alanine mutagenesis allows for an unbiased evaluation of the contributions of side chains to the folding and stability of a protein's structure (Yu et al., 1995). Each of the 103 buried, non-alanine residues was individually replaced with alanine and the mutant chains expressed in *E. coli*.

Capping regions exist in β-sheet structures to fulfill hydrogen bonding potential and prevent aggregation (Richardson and Richardson, 2002). To investigate this in the case of the β-helix, all residues in the N-terminal capping region (122-146) and the C-terminal region of the β-helix domain (521- 545) were chosen as targets of

Sheet:	Turn/Loop	A	A	A	A	Turn	B	B	B	B	Turn/Loop	Turn/Loop	Loop	Loop	Loop	C	C	C	C	C	Turn/Loop	Turn/Loop	Turn/Loop
Stack:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1			G	L	L I	D	R	D	Y	N		F	Y	G	G		E	T	V	D		F	GG K
2	V	L	T	I	E C	K	A	K	F	I		G	D	G	N		L	I	F	T	K	L	G
3	KG	S	R	I	A	G	V	F	M	E	Dorsal Fin	S			T		L	E	I	R	E	C	I
4	G	V	E	V	H	R	A	S	G	L	M	A			G		F	L	F	R	G	C	H
5	F	C	K	M	V	D	A	NN	P	S	G	G	K	D	G	I	I	T	F	E		N	Loop
6	G	N	Y	V	I	G	G	R	T	S	Y	G	S	V	S	S	A	Q	F	L	R	N	N
7	Loop	G	G	V	I	G	F	T	S	Y	R	A	G	E	S	G	V	K	T	W	Loop	N	Y
8	N	L	Q	F	R	D	S	V	V	I	Y	P	V	W	D	G	F	D	L	G	Ventral Fin	L	
9	N	H	L	I	D	N	L	L	V	R	G	A	L	G	V	G	F	G	M	D		G	K
10	G	M	Y	V	S	N	I	T	V	E	D	C	A	G	S	G	A	Y	L	L		T	H
11	E	S	V	F	T	N	I	A	I	I	D	T	Loop			Q	I	Y	I	S	G		
12	A	C	R	V	N	G	L	R	L	Loop		G	Loop			T	I	D	A	P	N		
13		S	T	V	S	G	I	T	G	M		V	D			P S R	I	N	V				
14		A	N	L																			

Figure 2-1. Structure-based sequence alignment of the P22 tailspike β-helix. The amino acid sequence of the β-helix domain (143 – 540) was aligned based on the stacking observed in the PDB structure 1TYU (Steinbacher et al., 1996). Each row in the above table represents a rung of the β-helix, as numbered on the left. Inward-pointing, buried core stack residues are colored according to the scheme used in Figure 1-9, while outward pointing residues and loop regions are not shaded. G511, which is disordered in the crystal structure, is included in the diagram as a potential part of stack 12, but is not colored.

mutagenesis, irrespective of their solvent accessibility. Six further solvent exposed residues located throughout the β-helix domain were added to the target list due to their proximity and potential chemical reactivity with cysteine residues. One disordered glycine, G511, was chosen for mutagenesis due to its potential location at a stacking position.

As further positional controls, five residues outside of the β-helix domain in a part of the structure known as the cysteine annulus were chosen for mutagenesis. The cysteine annulus is known to be important for the oligomerization of P22 tailspike (Danek and Robinson, 2003). The positions selected included two cysteine residues (C613 and C635) believed to be important for the presence of a transient disulfide bond in a oligomerized protein folding intermediate as well as residues that appeared likely to affect the reactivity of those cysteine residues (H610, D632, and P638) (Fomenko and Gladyshev, 2003).

B. MATERIALS AND METHODS

1. Production of Single Alanine Mutants

Gene 9, which encodes the Salmonella phage P22's tailspike protein, was previously cloned into the vector pET11a (Novagen) as described (Kreisberg et al., 2002). The resulting vector, pET(*gene9*), encoded the full length wild-type tailspike protein.

Methylated pET(*gene9*) DNA was purified to act as template DNA for mutagenesis from XL1-Blue cells (Stratagene) using a QIAGEN Plasmid Midi Kit. Scanning alanine mutagenesis was performed using the PCR-based QuikChange mutagenesis procedure (Stratagene) in a parallelized, high-throughput manner. Sequences of mutagenic primers (IDT) are listed in Table 2-1. Mutagenesis reactions were performed in parallel in 96 well plates and utilized multi-channel pipeting to achieve high-throughput processing.

Table 2-1. Mutagenic primers used to construct single alanine mutants. Forward mutagenic primers that were used to produce the specified single alanine mutations are listed. Reverse mutagenic primers used in the reaction were the exact reverse complement of the forward primers.

Mutation	Forward primer
K122A	5'-CAATAGAAGCTGATAAAAAATTTGCGTATTCAGTAAAATTATCAG-3'
Y123A	5'-CAATAGAAGCTGATAAAAAATTTAAGGCTTCAGTAAAATTATCAG-3'
S124A	5'-CAATAGAAGCTGATAAAAAATTTAAGTATGCAGTAAAATTATCAG-3'
V125A	5'-GCTGATAAAAAATTTAAGTATTCAGCAAAATTATCAGATTATCC-3'
K126A	5'-GTATTCAGTAGCATTATCAGATTATCCAACATTGCAGGATGCAG-3'
L127A	5'-GTATTCAGTAAAAGCATCAGATTATCCAACATTGCAGGATGCAGC-3'
S128A	5'-CAGTAAAATTAGCAGATTATCCAACATTGCAGGATGCAGC-3'
D129A	5'-CAGTAAAATTATCAGCTTATCCAACATTGCAGGATGCAGC-3'
Y130A	5'-GTAAAATTATCAGATGCTCCAACATTGCAGGATGCAGCATCTG-3'
P131A	5'-GTAAAATTATCAGATTATGCAACATTGCAGGATGCAGCATCTG-3'
T132A	5'-CAGATTATCCAGCATTGCAGGATGCAGCATCTGC-3'
L133A	5'-GATTATCCAACAGCGCAGGATGCAGCATCTGCTGC-3'
Q134A	5'-GATTATCCAACATTGGCGGATGCAGCATCTGCTGCG-3'
D135A	5'-CAACATTGCAGGCTGCAGCATCTGCTGCG-3'
S138A	5'-GCAGGATGCAGCAGCTGCTGCGGTTG-3'
V141A	5'-CAGCATCTGCTGCGGCTGATGGCCTTC-3'
D142A	5'-CTGCTGCGGTTGCTGGCCTTCTTATCGATC-3'
G143A	5'-GCATCTGCTGCGGTTGATGCCCTTCTTATCG-3'
L144A	5'-CTGCTGCGGTTGATGGCGCTCTTATCGATCGAG-3'
L145A	5'-CTGCTGCGGTTGATGGCCTTGCTATCGATCGAG-3'
I146A	5'-GATGGCCTTCTTGCCGATCGAGATTATAATTTTTATGGTGGAG-3'
R148A	5'-CGGTTGATGGCCTTCTTATCGATGCAGATTATAATTTTTATGGTG-3'
Y150A	5'-GGCCTTCTTATCGATCGAGATGCTAATTTTTATGGTGGAGAG-3'
F152A	5'-GGCCTTCTTATCGATCGAGATTATAATGCTTATGGTGGAGAG-3'
Y153A	5'-GAGATTATAATTTTGCTGGTGGAGAGACAGTTGATTTTGGCGG-3'
E156A	5'-GATTATAATTTTTATGGTGGAGCGACAGTTGATTTTGGCGG-3'
V158A	5'-GGTGGAGAGACAGCTGATTTTGGCGGAAAGG-3'
F160A	5'-GGTGGAGAGACAGTTGATGCTGGCGGAAAGGTTC-3'

Mutation	Forward primer
K163A	5'-GAGAGACAGTTGATTTTGGCGGAGCGGTTCTGACTATAG-3'
L165A	5'-GAGACAGTTGATTTTGGCGGAAAGGTTGCGACTATAGAATG-3'
T166A	5'-GATTTTGGCGGAAAGGTTCTGGCTATAGAATGTAAAGC-3'
I167A	5'-GTTGATTTTGGCGGAAAGGTTCTGACTGCAGAATGTAAAGC-3'
C169A	5'-GCGGAAAGGTTCTGACTATAGAAGCTAAAGCTAAATTTATAGGAG-3'
K172A	5'-CGGAAAGGTTCTGACTATAGAATGTAAAGCTGCATTTATAGGAG-3'
F173A	5'-GGTTCTGACTATAGAATGTAAAGCTAAAGCTATAGGAGATGG-3'
G175A	5'-GCTAAATTTATAGCAGATGGAAATCTTATTTTTACGAAATTAGGC-3'
L179A	5'-GAGATGGAAATGCTATTTTTACGAAATTAGGCAAAGGTTCCCGC-3'
F181A	5'-GATGGAAATCTTATTGCTACGAAATTAGGCAAAGGTTCCCGC-3'
L184A	5'-CTTATTTTTACGAAAGCAGGCAAAGGTTCCCGCATTGCCG-3'
S188A	5'-GAAATTAGGCAAAGGTGCCCGCATTGCCGG-3'
I190A	5'-CAAAGGTTCCCGCGCTGCCGGGGTTTTTATGG-3'
V193A	5'-CCCGCATTGCCGGGGCTTTTATGGAAAGC-3'
M195A	5'-CCGCATTGCCGGGGTTTTTGCGGAAAGCACTAC-3'
S260A	5'-GCAAAAGGGCAAAACATAACGGCTACGTTAGAAATTAGAG-3'
L262A	5'-CATAACGTCTACGGCAGAAATTAGAGAATGTATAGGGGTCG-3'
I264A	5'-GTCTACGTTAGAAGCTAGAGAATGTATAGGGGTCGAAGTTC-3'
C267A	5'-GAAATTAGAGAAGCTATAGGGGTCGAAGTTCATCGGGCTAG-3'
V270A	5'-GAGAATGTATAGGGGCCGAAGTTCATCGGGC-3'
V272A	5'-GTATAGGGGTCGAAGCTCATCGGGCTAGC-3'
G277A	5'-GTTCATCGGGCTAGCGCTCTAATGGCTGG-3'
F282A	5'-GGCTAGCGGTCTAATGGCTGGTGCTTTGTTTAGAGG-3'
F284A	5'-GCGGTCTAATGGCTGGTTTTTTGGCTAGAGGGTGTC-3'
C287A	5'-GTTTAGAGGGGCTCACTTCTGCAAGATGGTAGACGC-3'
C290A	5'-GAGGGTGTCACTTCGCCAAGATGGTAGACGCC-3'
M292A	5'-CTTCTGCAAGGCGGTAGACGCCAATAATCCAAGCG-3'
V293A	5'-CTGCAAGATGGCAGACGCCAATAATCCAAGCG-3'
P298A	5'-GACGCCAATAATGCAAGCGGAGGTAAAGATGGC-3'
G300A	5'-GACGCCAATAATCCAAGCGCAGGTAAAGATGGC-3'
G301A	5'-GCCAATAATCCAAGCGGAGCTAAAGATGGCATTATAAC-3'
I305A	5'-GTAAAGATGGCGCTATAACCTTCGAAAACCTTAGCGGCG-3'

Mutation	Forward primer
I306A	5'-GTAAAGATGGCATTGCAACCTTCGAAAACCTTAGCGGCG-3'
F308A	5'-CATTATAACCGCCGAAAACCTTAGCGGCGATTGGGG-3'
N310A	5'-CATTATAACCTTCGAAGCCCTTAGCGGCGATTGGGG-3'
N319A	5'-GGGGAAGGGTGCCTATGTCATTGGCGGACG-3'
V321A	5'-GGGGAAGGGTAACTATGCCATTGGCGGAC-3'
G324A	5'-CTATGTCATTGGCGCACGAACCAGCTATGGG-3'
T326A	5'-CATTGGCGGACGAGCCAGCTATGGGTC-3'
G329A	5'-GAACCAGCTATGCGTCAGTAAGTAGCGCCC-3'
S333A	5'-CAGCTATGGGTCAGTAAGTGCCGCCCAGTTTTTACG-3'
F336A	5'-GGGTCAGTAAGTAGCGCCCAGGCTTTACGTAATAATGG-3'
N339A	5'-GCGCCCAGTTTTTACGTGCTAATGGTGGCTTTGAACG-3'
G347A	5'-GGTGGCTTTGAACGTGATGCTGGAGTTATTGGG-3'
V349A	5'-GCTTTGAACGTGATGGTGGAGCTATTGGGTTTAC-3'
F352A	5'-GGAGTTATTGGGGCTACTTCATATCGCGCTGGGG-3'
S354A	5'-GTTATTGGGTTTACTGCATATCGCGCTGGGGAG-3'
G361A	5'-CGCGCTGGGGAGAGTGCCGTTAAAACTTG-3'
V362A	5'-GCTGGGGAGAGTGGCGCTAAAACTTGGC-3'
T364A	5'-GTGGCGTTAAAGCTTGGCAAGGTACTGTGGG-3'
N376A	5'-GGGCTCGACAACCTCTCGCGCCTATAATCTGC-3'
L379A	5'-CGCAACTATAATGCGCAATTCCGCGACTCGGTCG-3'
F381A	5'-CTCGCAACTATAATCTGCAAGCCCGCGACTCGG-3'
S384A	5'-CAATTCCGCGACGCGGTCGTTATTTACCCC-3'
V386A	5'-CCGCGACTCGGTCGCTATTTACCCCG-3'
P389A	5'-CTCGGTCGTTATTTACGCCGTATGGGACGG-3'
G393A	5'-CCCCGTATGGGACGCATTCGATTTAGGTGC-3'
F394A	5'-CCCCGTATGGGACGGAGCCGATTTAGGTGC-3'
L396A	5'-GGGACGGATTCGATGCAGGTGCTGACACTGAC-3'
L424A	5'-CAATACCCACTGCATCAGTTACCCGCAAATCACCTGATTG-3'
H426A	5'-GTTACCCCTAAATGCCCTGATTGATAATCTTCTGGTTCGC-3'
I428A	5'-CTAAATCACCTGGCTGATAATCTTCTGGTTCGCGGGG-3'
L431A	5'-CACCTGATTGATAATGCTCTGGTTCGCGGGGCG-3'
V433A	5'-GATAATCTTCTGGCTCGCGGGGCGTTAGG-3'

Mutation	Forward primer
G440A	5'-GCGGGGCGTTAGGTGTAGCTTTTGGTATGG-3'
F441A	5'-CGGGGCGTTAGGTGTAGGTGCTGGTATGGATG-3'
M443A	5'-GCGTTAGGTGTAGGTTTTGGTGCGGATGGTAAGGG-3'
G445A	5'-GGTTTTGGTATGGATGCTAAGGGCATGTATGTGTC-3'
M448A	5'-GATGGTAAGGGCGCGTATGTGTCTAATATTACCGTAGAAG-3'
V450A	5'-GGATGGTAAGGGCATGTATGCGTCTAATATTACCG-3'
I453A	5'-GTATGTGTCTAATGCTACCGTAGAAGATTGCGCTGGGTC-3'
V455A	5'-CTAATATTACCGCAGAAGATTGCGCTGGGTCTGG-3'
C458A	5'-CGTAGAAGATGCCGCTGGGTCTGGCGCG-3'
S461A	5'-GATTGCGCTGGGGCTGGCGCGTACC-3'
G462A	5'-CGCTGGGTCTGCCGCGTACCTACTC-3'
L465A	5'-GGGTCTGGCGCGTACGCACTCACCCACG-3'
T467A	5'-GCGCGTACCTACTCGCCCACGAATCAG-3'
S470A	5'-CCTACTCACCCACGAAGCAGTATTTACCAATATAGC-3'
F472A	5'-GCGTACCTACTCACCCACGAATCAGTAGCTACCAATATAG-3'
I475A	5'-CCCACGAATCAGTATTTACCAATGCAGCCATAATTGACACC-3'
I477A	5'-CAATATAGCCGCAATTGACACCAATACTAAGGATTTCCAGGC-3'
T480A	5'-GCCATAATTGACGCCAATACTAAGGATTTCCAGGC-3'
Q489A	5'-CCAGGCGAATGCGATTTATATATCTGGGGCTTGCCG-3'
I490A	5'-GGCGAATCAGGCTTATATATCTGGGGCTTGCCGTG-3'
I492A	5'-GGCGAATCAGATTTATGCATCTGGGGCTTGCCGTG-3'
S493A	5'-GATTTATATAGCTGGGGCTTGCCGTGTGAACGG-3'
C496A	5'-GATTTATATATCTGGGGCTGCCCGTGTGAACGGTTTACG-3'
V498A	5'-GGGGCTTGCCGTGCGAACGGTTTACG-3'
N499A	5'-GGGGCTTGCCGTGTGGCCGGTTTACGTTTAATTG-3'
L501A	5'-CCGTGTGAACGGTGCACGTTTAATTGGGATCCGC-3'
L503A	5'-CCGTGTGAACGGTTTACGTGCAATTGGGATCCGC-3'
G511A	5'-CCGCTCAACCGATGCGCAGGGTCTAAC-3'
T515A	5'-GGCAGGGTCTAGCCATAGACGCCCC-3'
I516A	5'-GGCAGGGTCTAACCGCAGACGCCCCTAAC-3'
S521A	5'-GACGCCCCTAACGCTACCGTAAGCGG-3'
T522A	5'-CCCTAACTCTGCCGTAAGCGGTATAACCGG-3'

Mutation	Forward primer
V523A	5'-CCCCTAACTCTACCGCAAGCGGTATAACCG-3'
S524A	5'-CCCCTAACTCTACCGTAGCCGGTATAACCGGG-3'
G525A	5'-CTACCGTAAGCGCTATAACCGGGATGGTAGAC-3'
I526A	5'-CGTAAGCGGTGCAACCGGGATGGTAGACCC-3'
T527A	5'-GTAAGCGGTATAGCCGGGATGGTAGACCC-3'
G528A	5'-GCGGTATAACCGCGATGGTAGACCCCTC-3'
M529A	5'-GCGGTATAACCGGGGCGGTAGACCCCTC-3'
V530A	5'-GGTATAACCGGGATGGCAGACCCCTCTAG-3'
D531A	5'-CGGGATGGTAGCCCCCTCTAGAATTAATGTTGC-3'
P532A	5'-CCGGGATGGTAGACGCCTCTAGAATTAATGTTGC-3'
S533A	5'-CCGGGATGGTAGACCCCGCTAGAATTAATGTTG-3'
R534A	5'-GTATAACCGGGATGGTAGACCCCTCTGCAATTAATGTTGC-3'
I535A	5'-CCCCTCTAGAGCTAATGTTGCTAATTTGGCAGAAGAAGGG-3'
N536A	5'-GGGATGGTAGACCCCTCTAGAATTGCTGTTGCTAATTTGG-3'
V537A	5'-GGTAGACCCCTCTAGAATTAATGCTGCTAATTTGGC-3'
N539A	5'-CCCCTCTAGAATTAATGTTGCTGCTTTGGCAGAAGAAGGG-3'
L540A	5'-GACCCCTCTAGAATTAATGTTGCTAATGCGGCAGAAGAAG-3'
E542A	5'-CTAATTTGGCAGCAGAAGGGTTAGGTAATATCCGC-3'
E543A	5'-CTAATTTGGCAGAAGCAGGGTTAGGTAATATCCGC-3'
G544A	5'-GGCAGAAGAAGCGTTAGGTAATATCCGCGCTAATAG-3'
L545A	5'-GGCAGAAGAAGGGGCAGGTAATATCCGCGCTAATAG-3'
H610A	5'-GCTGTATCATTAAAAGTTAACGCCAAAGATTGCAGGGGGG-3'
C613A	5'-CCACAAAGATGCCAGGGGGGCAGAGATACCATTTG-3'
D632A	5'-CGCGTCAGATGATTTTATAAAGGCTTCCTCATGTTTTTTGC-3'
C635A	5'-CAGATGATTTTATAAAGGATTCCTCAGCTTTTTTGCCATATTGGG-3'
P638A	5'-GGATTCCTCATGTTTTTTGGCATATTGGGAAAATAATTCTACTTC-3'

Clones encoding full length, but potentially altered, tailspike were identified via over-expression in *Escherichia coli* NovaBlue(DE3) cells (Novagen). Plasmid DNA was purified using QIAprep Spin MiniPrep kits (QIAGEN). DNA sequencing of a 400 base pair region (20% of the tailspike gene) containing the targeted codon confirmed mutant clones (MGH DNA Core, Cambridge, MA).

2. Expression of Tailspike Constructs

Plasmid DNA encoding confirmed alanine mutants was transformed into *E. coli* BL21(DE3) using a single step protocol, as described in (Chung et al., 1989). Transformations were performed in parallel in 2 ml 96 deep-well plates (Bel-Art) and were prepared in a high-throughput manner. Cultures were selected on individual Luria Broth (LB) plates containing 100 μg/ml ampicillin. This selection stage was the only part of the experimental procedure that could not be performed in a high-throughput way.

Cultures were made into 40% glycerol working stocks and 8.3% DMSO freeze stocks after 8 hours of growth in LB containing 100 μg/ml ampicilllin at 37ºC. Working stocks were stored at -20ºC in 96 deep-well plates while freeze stocks were stored at -80ºC in dram vials. Expression cultures were inoculated via 100-fold dilution of the working stocks into LB media containing 0.5% glucose, 50 mM 3-[N-morpholino]-propanesulfonic acid (MOPS) and 100 μg/ml ampicillin. Cells were grown shaking (225 rpm) in 500 μl volumes in 96 deep-well plates for approximately 4 hours at 30ºC until cultures reached an OD_{600} between 0.4 and 0.5. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1.5 mM to induce protein expression. Cultures were transferred to 37ºC, 30ºC, 18ºC, or 10ºC and were allowed to express protein for approximately 4.25, 6.5, 18.3 hours, or 24 hours respectively. Cultures were harvested via centrifugation for 15 minutes at 4000*g* and 4ºC. Cells were resuspended to 1/10 volume in lysis buffer (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA; 0.1% Triton X-100) and stored at -20ºC.

3. Cell Lysis and SDS-PAGE

Frozen expression samples were thawed at room temperature and lysis was initiated by addition of freshly dissolved lysozyme (1 mg/ml final concentration from 10 mg/ml aqueous solution) and PMSF (9 mM final concentration from 90 mM stock in isopropanol). After 30 min of slow stirring at room temperature and addition of DNase I (Sigma) (10 μg/ml final concentration from 0.06 mg/ml stock in glycerol; 1 M $MgSO_4$), lysates were stirred for an additional 30 min.

Half of each lysate was removed as a complete lysate sample, while the remaining lysate was centrifuged for 30 min at 4,000*g* and 4ºC in shallow well plates to separate soluble and insoluble fractions. Lysate pellets were resuspended in lysis buffer by gentle agitation. Lysate fractions were mixed with 2% SDS sample buffer containing 15% BME as previously described (Betts et al., 2004). SDS-PAGE was run at 4ºC on 7.5% polyacrylamide, 102-lane Triple Wide gels (C.B.S. Scientific) using the discontinuous buffer system of King and Laemmli (King and Laemmli, 1971). A multichannel pipet was used to ensure rapid and efficient loading.

4. Optical Densitometry and Folding Efficiency Calculations

Coomassie-stained gels were individually scanned at 300ppi and 8-bit gray scale via a ScanMaker 9800XL tabletop scanner (Microtek). Optical densitometry of the gel image was performed using the GelAnalyzer plug-in (Appendix A) for ImageJ (NIH). In brief, gel images were calibrated using a Kodak OD step tablet (AI-04-0057, Tiffen Company) that ranged from 0.6 to 3.09 OD. Profile plots of lanes were calculated as the average OD across a lane's width. Background levels of cellular protein expression, as calculated as the average profile plot of expression cultures containing pET11a vectors lacking any tailspike gene, were scaled and subtracted from each complete lysate sample lane. Integration of the positive area of peaks corresponding to tailspike bands, as determined from the migration of purified tailspike standards, yielded relative protein amounts.

Folding efficiencies were calculated as the percent of all tailspike chains found in the native, trimeric tailspike (NT) band, according to the following formula:

$$\text{Percent Folded} = \frac{OD_{NT}}{OD_{NT} + OD_{U}} * 100\%$$

Folding efficiencies were normalized to the average wild-type folding efficiency, which was calculated as the average of 22 independent lysates, according to the formula:

$$\text{Folding Efficiency} = \frac{\text{Percent Folded}_{\text{Mutant}}}{\langle \text{Percent Folded}_{\text{WT}} \rangle} * 100\%$$

C. RESULTS

1. Production of Single Alanine Mutants

In total, 150 non-alanine sites were individually mutated to alanine, as confirmed by DNA sequencing. As detailed in Section B1, DNA sequencing was performed on a 400 base pair region containing the targeted codon. Although the entire gene was not sequenced for each individual mutant, only one sample out of 241 samples sequenced had a second site mutation outside the primer region, resulting in approximately 0.02 second site mutations per gene. This low error rate indicated that this limited sequencing was sufficient to confirm single alanine mutations, thus allowing for these mutants to be assayed for their affect on folding, as detailed below.

In addition to the desired 150 single alanine mutants, a small number of insertion or deletion mutants were produced. The deletion of base T553 in the codon encoding L184 of tailspike or deletion of base A586 in the M195-encoding codon resulted in reading frame shifts that produced truncated proteins 214 amino acids in length that appeared exclusively in the insoluble fraction of 37ºC expression lysates. Another deletion mutant, a deletion of T1564 in the codon encoding S521 generated a frame shift that produced a 523 amino acid polypeptide chain. When expressed at 37ºC, this protein also appeared exclusively in the insoluble fraction of the cellular lysate.

An insertional mutant was recovered that contained an F472A mutation with a subsequent duplication of residues 468-472, thus creating an insertion of 6 amino acids into the tailspike protein that included two F→A mutations. When expressed at 37ºC, 30ºC, or 18ºC, this insertional duplication yielded insoluble, SDS-sensitive polypeptide chains.

2. *In Vivo* Expression of Single Alanine Mutants

All single alanine mutants were individually over-expressed from a pET plasmid in *Escherichia coli* cultures incubated at 37ºC, 30ºC, and 18ºC in order to assess the contribution of the altered residue to the *in vivo* folding of the β-helix domain. Some mutant strains were sensitive to freeze stock storage conditions. These strains would not express tailspike proteins even after several rounds of growth following freeze storage and would only produce tailspike protein after the appropriate plasmids were freshly transformed into bacteria.

Cells containing pET11a plasmids lacking any tailspike genes were induced as a negative control of tailspike expression. Positive controls included wild-type and ΔN tailspike, which lacks the N-terminal domain (1-108) (Miller et al., 1998b). Mutant expression at each temperature was performed in triplicate. Induced samples were lysed, mixed with SDS sample buffer, and unheated samples were analyzed by SDS-PAGE. Sample gels are shown in Figure 2-2.

As can be seen in Figure 2-2, host protein translational levels were consistent across all samples. Mutant and wild-type lysates accumulated tailspike polypeptide chains as clearly resolved bands. These observations proved true for all expression samples at all three temperatures. Thus, alanine substitutions at the targeted sites had no noticeable affect on translation or degradation of the tailspike polypeptide chain.

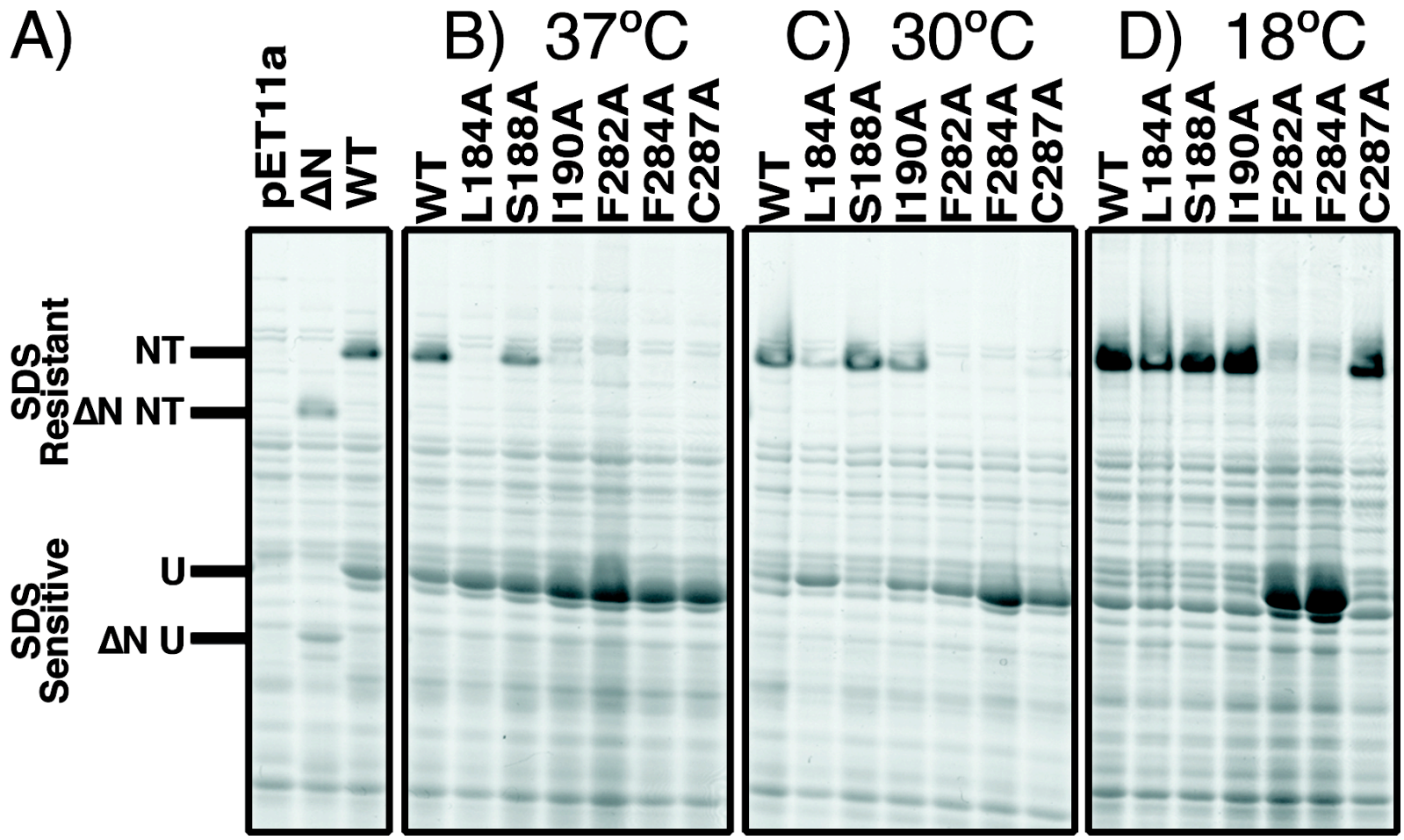

Figure 2-2. Sample SDS-PAGE data of complete cellular lysates. (A) Positive (WT & ΔN) and negative (pET11a) control samples expressed at 37ºC. WT and ΔN expression samples produce some native, SDS-resistant trimers (NT) as well as non-native, SDS-sensitive chains that unfold (U) upon mixture with SDS. WT and mutant samples are shown for expression at (B) 37ºC, (C) 30ºC, and (D) 18ºC. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

When complete *E. coli* lysate samples were electrophoresed through an SDS polyacrylamide gel, over-expressed tailspike chains partitioned between two bands in the gel (Figure 2-2). Chains that reached the SDS-resistant, native tailspike structure did not dissociate in the presence of SDS (Goldenberg et al., 1982). These chains appeared on the gel as a slowly migrating native trimer band (NT) of approximately 210-kD. In contrast, partially folded intermediates and inclusion body aggregates are dissociated and unfolded in detergent, causing these species to electrophorese as a rapidly migrating, 72-kD band (U). Quantification of the percent of tailspike polypeptide chains reaching the native state provided an accurate and precise measure of the *in vivo* partitioning between productive and non-productive folding.

Negative control samples were used to subtract background cellular protein levels so that tailspike bands could be accurately quantified. Normalization of the percent folded measurement to the average wild-type value generated a useful measure, the percent of wild-type folding efficiency. Comparison of these folding efficiencies allowed for an unbiased assessment of the effects of individual alanine substitutions with respect to a known standard, the wild-type sequence.

3. Overall Folding Efficiencies

At each temperature, the distribution of folding efficiencies among mutants proved bimodal, allowing for the classification of each mutant as either a wild-type-like folder or as folding deficient. Mutants whose folding efficiencies were less than 75% of wild-type's folding efficiency were classified as folding deficient. This 75% threshold was conservatively chosen based upon the distribution of averages, examination of the statistical overlap of sample values with wild-type, and the expectation that a mutant classified as defective at a low temperature should be similarly classified at higher temperatures.

Quantitative analysis of the *in vivo* folding of all mutants at all three temperatures is shown in Figure 2-3 and Table 2-2. Though a few mutant sequences were able to reach the native state at the presumed physiological temperature of the *Salmonella* phage P22,

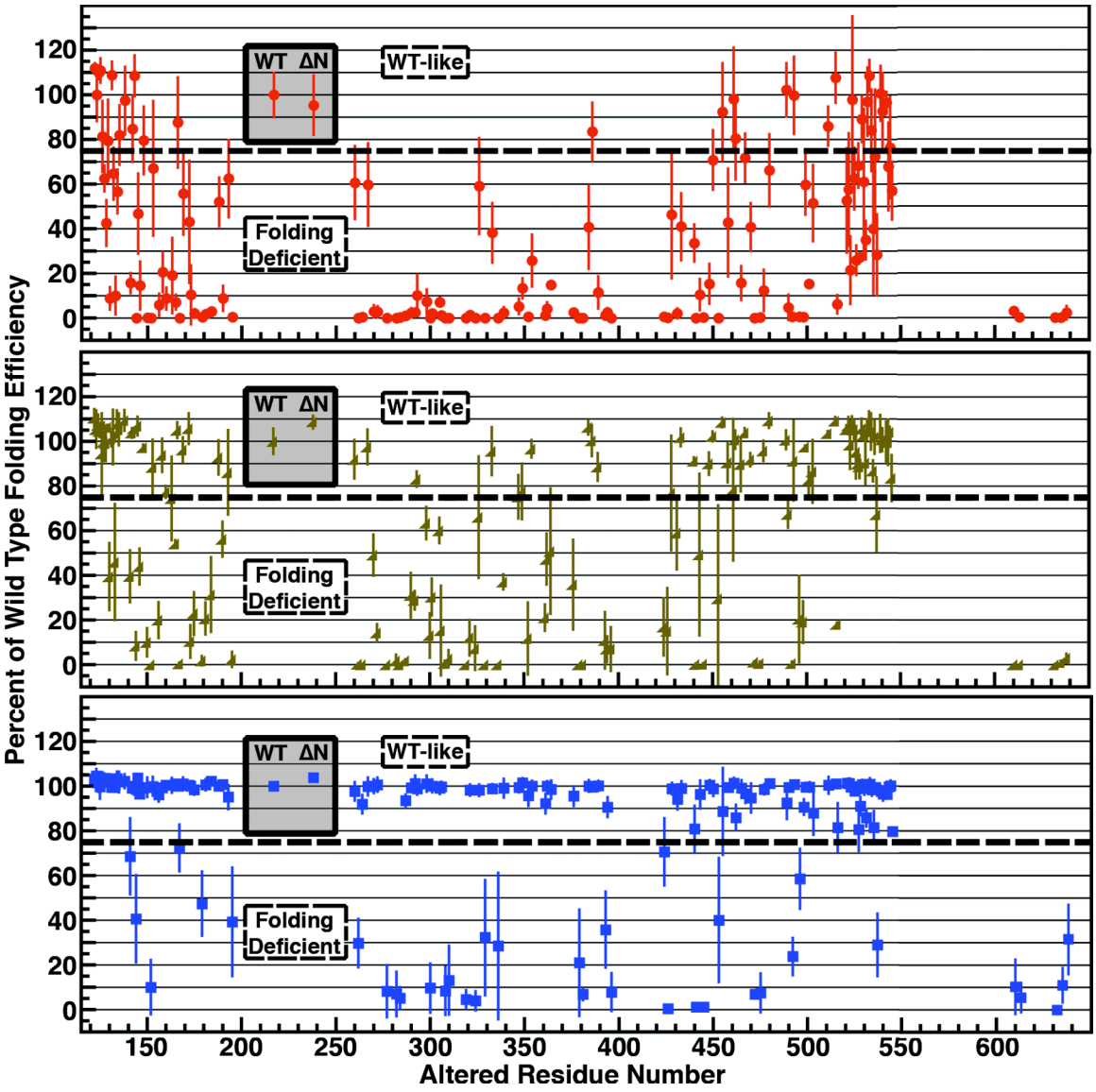

Figure 2-3. Quantitative results of folding experiments. Average percent of wild-type folding efficiencies are shown for all alanine mutants when expressed at 37ºC (red circle), 30ºC (brown triangle), and 18ºC (blue square). Colored lines indicate the standard deviations of triplicate experiments. Note that WT and ΔN average percent of wild-type folding efficiencies and standard deviations are shown as insets. The 75% threshold delineating wild-type-like folders and folding deficient samples is shown as a dashed line.

Table 2-2. Folding efficiencies of single alanine mutants. The percent of WT folding efficiencies for all 150 single alanine substitutions, as well as the WT and ΔN positive controls, at the four temperatures tested are given as averages and standard deviations of multiple expression experiments, as described in this chapter. 10ºC expression data was not collected for mutants capable of folding like wild-type at 18ºC. These samples are marked by "NA". Folding deficient average values (<75%) are shaded and bolded.

Sample	Percent of WT folding efficiency			
	37ºC	30ºC	18ºC	10ºC
WT	100.0 ± 9.9	100.0 ± 5.7	100.0 ± 1.8	100.0±4.3
ΔN	95.4 ± 13.3	108.9 ± 3.2	103.9 ± 1.0	100.5±4.4
K122A	111.9 ± 2.3	110.4 ± 4.1	103.3 ± 3.0	NA
Y123A	100.1 ± 11.8	105.4 ± 8.4	104.6 ± 3.0	NA
S124A	109.9 ± 2.9	107.3 ± 4.7	101.5 ± 4.0	NA
V125A	110.9 ± 5.4	104.6 ± 2.6	99.8 ± 5.3	NA
K126A	81.4 ± 15.9	94.4 ± 18.1	102.3 ± 4.1	NA
L127A	**62.6 ± 5.8**	99.8 ± 8.6	101.8 ± 4.0	NA
S128A	**42.6 ± 10.2**	99.1 ± 7.8	103.6 ± 1.2	NA
D129A	79.4 ± 18.4	106.3 ± 1.1	103.2 ± 2.0	NA
Y130A	**8.9 ± 5.0**	**39.4 ± 15.0**	102.4 ± 0.8	NA
P131A	108.8 ± 6.1	101.2 ± 3.3	99.6 ± 1.7	NA
T132A	**64.8 ± 11.6**	102.0 ± 12.2	99.5 ± 1.0	NA
L133A	**10.1 ± 8.5**	**46.0 ± 25.9**	100.4 ± 0.5	NA
Q134A	**56.7 ± 9.8**	107.9 ± 6.0	103.4 ± 3.2	NA
D135A	82.0 ± 13.3	105.4 ± 7.2	101.9 ± 4.7	NA
S138A	97.5 ± 15.1	109.4 ± 4.6	102.2 ± 3.1	NA
V141A	**15.7 ± 4.5**	**39.6 ± 11.6**	**68.6 ± 17.0**	**40.6±12.1**
D142A	84.8 ± 14.8	103.8 ± 2.2	99.2 ± 2.4	NA
G143A	108.5 ± 9.1	105.7 ± 0.4	100.1 ± 3.9	NA
L144A	**0.0 ± 0.0**	**8.4 ± 6.2**	**40.7 ± 19.5**	**36.4±23.5**
L145A	**46.8 ± 17.9**	106.9 ± 4.2	103.7 ± 0.5	NA
I146A	**14.6 ± 10.7**	**44.0 ± 8.0**	96.6 ± 0.7	NA
R148A	79.6 ± 15.1	97.3 ± 1.2	99.2 ± 1.9	NA
Y150A	**0.2 ± 0.1**	**10.0 ± 6.3**	99.5 ± 4.4	NA
F152A	**0.0 ± 0.0**	**0.0 ± 0.0**	**10.1 ± 12.1**	**7.7±6.7**

Sample	37°C	30°C	18°C	10°C
Y153A	**67.1 ± 30.2**	88.4 ± 12.4	99.9 ± 4.2	NA
E156A	**6.0 ± 4.9**	**20.0 ± 8.0**	96.2 ± 3.1	NA
V158A	**20.6 ± 9.0**	93.8 ± 7.4	99.3 ± 1.6	NA
F160A	**8.8 ± 4.9**	77.4 ± 2.7	100.4 ± 0.8	NA
K163A	**19.1 ± 16.9**	**74.4 ± 18.7**	100.4 ± 3.3	NA
L165A	**7.1 ± 3.3**	**54.2 ± 1.5**	101.0 ± 0.2	NA
T166A	87.6 ± 20.2	105.0 ± 3.5	100.1 ± 1.1	NA
I167A	**0.0 ± 0.0**	**0.5 ± 0.4**	**72.4 ± 10.4**	**68.5±19.9**
C169A	**55.7 ± 18.3**	96.2 ± 5.5	101.0 ± 2.9	NA
K172A	**43.1 ± 27.4**	105.7 ± 6.8	100.4 ± 1.1	NA
F173A	**10.4 ± 13.1**	**10.4 ± 7.2**	100.0 ± 1.9	NA
G175A	**2.1 ± 1.9**	**22.9 ± 9.5**	98.4 ± 2.2	NA
L179A	**0.5 ± 0.4**	**1.8 ± 2.0**	**47.4 ± 14.3**	**68.9±16.4**
F181A	**1.8 ± 1.4**	**20.7 ± 7.1**	100.6 ± 2.5	NA
L184A	**3.0 ± 1.3**	**31.4 ± 16.7**	102.2 ± 1.5	NA
S188A	**52.1 ± 10.8**	92.8 ± 7.7	100.1 ± 1.4	NA
I190A	**8.9 ± 5.6**	**56.2 ± 7.8**	100.8 ± 0.4	NA
V193A	**62.5 ± 17.3**	86.1 ± 18.9	95.1 ± 5.3	NA
M195A	**0.5 ± 0.6**	**2.4 ± 3.4**	**39.3 ± 24.3**	**48.7±25.5**
S260A	**60.7 ± 16.3**	92.0 ± 8.6	97.7 ± 4.1	NA
L262A	**0.0 ± 0.0**	**0.3 ± 0.4**	**29.8 ± 10.8**	**28.5±21.6**
I264A	**0.4 ± 0.6**	**0.1 ± 0.1**	92.0 ± 4.1	NA
C267A	**59.8 ± 18.4**	97.5 ± 7.8	99.9 ± 3.9	NA
V270A	**3.0 ± 2.7**	**49.1 ± 9.2**	100.0 ± 3.1	NA
V272A	**2.8 ± 2.5**	**14.3 ± 3.8**	100.7 ± 3.2	NA
G277A	**0.0 ± 0.0**	**0.0 ± 0.0**	**8.3 ± 11.6**	**0.0±0.0**
F282A	**0.0 ± 0.0**	**1.8 ± 2.6**	**7.0 ± 9.9**	**1.6±2.2**
F284A	**0.3 ± 0.4**	**0.0 ± 0.0**	**5.2 ± 4.8**	**4.1±2.9**
C287A	**1.1 ± 1.3**	**1.7 ± 0.7**	93.6 ± 2.6	NA
C290A	**2.4 ± 2.1**	**31.1 ± 10.0**	99.3 ± 1.1	NA
M292A	**2.5 ± 1.8**	**28.8 ± 4.1**	101.3 ± 3.5	NA
V293A	**10.2 ± 9.2**	83.1 ± 3.4	98.9 ± 2.7	NA

Sample	37°C	30°C	18°C	10°C
P298A	**7.4 ± 5.4**	**63.4 ± 7.2**	101.1 ± 3.6	NA
G300A	**0.0 ± 0.0**	**13.1 ± 9.9**	**9.7 ± 11.0**	**26.3±20.3**
G301A	**2.1 ± 1.5**	**30.3 ± 8.2**	100.1 ± 2.9	NA
I305A	**7.1 ± 2.0**	**60.1 ± 5.7**	99.1 ± 2.5	NA
I306A	**1.2 ± 0.9**	**15.3 ± 20.1**	99.8 ± 2.2	NA
F308A	**0.0 ± 0.0**	**0.7 ± 1.0**	**8.4 ± 10.7**	**3.5±4.4**
N310A	**0.0 ± 0.0**	**2.7 ± 3.9**	**13.2 ± 15.4**	**15.8±18.2**
N319A	**0.0 ± 0.0**	**0.0 ± 0.0**	**4.6 ± 4.1**	**6.0±5.4**
V321A	**1.4 ± 1.6**	**12.1 ± 7.9**	98.3 ± 2.5	NA
G324A	**0.0 ± 0.1**	**7.3 ± 9.7**	**4.0 ± 4.2**	**6.3±5.1**
T326A	**59.1 ± 21.5**	**66.1 ± 27.2**	98.2 ± 2.6	NA
G329A	**0.0 ± 0.0**	**0.0 ± 0.0**	**32.3 ± 25.7**	**30.7±18.2**
S333A	**38.2 ± 13.3**	95.6 ± 10.7	98.9 ± 2.0	NA
F336A	**0.0 ± 0.0**	**0.1 ± 0.1**	**28.5 ± 32.8**	**42.3±27.3**
N339A	**2.4 ± 2.5**	**37.1 ± 3.4**	99.1 ± 4.2	NA
G347A	**5.2 ± 3.5**	75.2 ± 9.7	99.4 ± 2.2	NA
V349A	**13.4 ± 4.5**	77.6 ± 12.4	101.5 ± 2.0	NA
F352A	**0.6 ± 0.9**	**11.7 ± 16.1**	95.8 ± 4.5	NA
S354A	**25.8 ± 11.6**	96.7 ± 3.9	100.0 ± 2.3	NA
G361A	**1.3 ± 0.9**	**21.0 ± 5.9**	92.2 ± 4.4	NA
V362A	**4.2 ± 2.8**	**47.3 ± 11.4**	100.0 ± 2.2	NA
T364A	**14.9 ± 1.1**	**50.8 ± 28.0**	98.5 ± 4.1	NA
N376A	**2.5 ± 1.1**	**35.9 ± 20.1**	95.6 ± 4.3	NA
L379A	**0.0 ± 0.0**	**0.0 ± 0.0**	**21.0 ± 23.7**	**8.7±8.8**
F381A	**0.0 ± 0.0**	**0.1 ± 0.0**	**6.9 ± 2.6**	**15.6±13.3**
S384A	**40.8 ± 18.7**	106.3 ± 3.1	100.1 ± 2.6	NA
V386A	83.5 ± 13.1	100.3 ± 6.9	99.6 ± 0.9	NA
P389A	**11.5 ± 7.2**	88.5 ± 6.2	100.2 ± 2.2	NA
G393A	**1.2 ± 1.4**	**10.9 ± 12.7**	**35.8 ± 17.0**	**12.5±11.7**
F394A	**2.5 ± 2.0**	**7.0 ± 5.7**	90.6 ± 4.4	NA
L396A	**0.1 ± 0.2**	**7.0 ± 9.8**	**7.9 ± 8.5**	**3.3±2.3**
L424A	**0.7 ± 0.9**	**16.9 ± 12.8**	**70.6 ± 15.0**	**45.8±24.8**

Sample	37°C	30°C	18°C	10°C
H426A	**0.2 ± 0.3**	**15.0 ± 19.3**	**0.5 ± 0.8**	**2.5±3.3**
I428A	**46.4 ± 28.5**	76.9 ± 25.7	98.9 ± 2.6	NA
L431A	**2.1 ± 2.2**	**59.0 ± 16.3**	94.1 ± 4.5	NA
V433A	**41.0 ± 15.0**	101.8 ± 4.0	99.0 ± 3.2	NA
G440A	**33.6 ± 8.3**	91.3 ± 1.5	80.9 ± 10.3	NA
F441A	**0.1 ± 0.1**	**0.2 ± 0.3**	**1.2 ± 1.6**	**0.2±0.2**
M443A	**10.5 ± 7.0**	**49.3 ± 36.2**	96.3 ± 6.3	NA
G445A	**0.3 ± 0.4**	**0.3 ± 0.3**	**1.2 ± 1.7**	**0.4±0.6**
M448A	**15.4 ± 8.9**	90.0 ± 5.0	100.5 ± 2.2	NA
V450A	**70.8 ± 13.3**	102.2 ± 2.2	98.9 ± 3.3	NA
I453A	**0.1 ± 0.1**	**29.6 ± 41.8**	**40.1 ± 27.7**	**7.4±5.5**
V455A	92.3 ± 21.9	108.5 ± 2.1	88.7 ± 19.4	NA
C458A	**42.8 ± 24.1**	90.5 ± 8.9	99.4 ± 1.9	NA
S461A	98.1 ± 23.2	78.4 ± 31.7	101.4 ± 2.3	NA
G462A	80.5 ± 18.4	100.8 ± 4.1	86.0 ± 5.6	NA
L465A	**15.7 ± 7.6**	89.7 ± 12.1	99.1 ± 3.6	NA
T467A	**71.8 ± 10.9**	103.9 ± 2.0	96.7 ± 4.2	NA
S470A	**40.8 ± 10.7**	91.8 ± 2.9	94.8 ± 6.5	NA
F472A	**0.1 ± 0.1**	**1.2 ± 0.3**	**6.9 ± 1.3**	**3.3±2.3**
I475A	**0.3 ± 0.3**	**0.6 ± 0.8**	**7.5 ± 8.7**	**0.9±1.3**
I477A	**12.4 ± 9.2**	95.8 ± 4.5	98.6 ± 2.3	NA
T480A	**66.2 ± 16.1**	109.3 ± 3.2	101.3 ± 0.8	NA
Q489A	102.2 ± 11.9	100.7 ± 4.2	92.6 ± 7.2	NA
I490A	**4.8 ± 5.6**	**67.4 ± 6.0**	99.4 ± 1.1	NA
I492A	**0.7 ± 0.6**	**0.7 ± 0.7**	**23.8 ± 8.3**	**2.9±2.2**
S493A	99.8 ± 17.2	91.3 ± 17.8	100.8 ± 1.6	NA
C496A	**0.7 ± 1.0**	**20.4 ± 19.4**	**58.5 ± 13.4**	**15.1±3.9**
V498A	**0.5 ± 0.7**	**19.1 ± 9.3**	90.7 ± 3.5	NA
N499A	**59.8 ± 13.5**	97.5 ± 0.6	99.6 ± 1.9	NA
L501A	**15.3 ± 1.1**	82.5 ± 6.2	99.6 ± 2.1	NA
L503A	**51.5 ± 17.1**	86.6 ± 14.1	88.0 ± 9.8	NA
G511A	86.0 ± 8.8	103.6 ± 1.0	100.4 ± 3.7	NA

Sample	37°C	30°C	18°C	10°C
T515A	107.7 ± 11.2	109.1 ± 1.8	101.0 ± 1.2	NA
I516A	**6.2 ± 4.1**	**18.1 ± 0.8**	81.5 ± 10.8	NA
S521A	**52.7 ± 23.4**	105.4 ± 1.8	101.4 ± 1.9	NA
T522A	**57.6 ± 25.2**	107.1 ± 3.8	101.2 ± 2.2	NA
V523A	**21.5 ± 15.2**	98.4 ± 10.7	99.6 ± 0.8	NA
S524A	97.8 ± 37.3	107.3 ± 4.0	99.9 ± 1.1	NA
G525A	**62.5 ± 13.8**	105.1 ± 4.5	97.8 ± 1.5	NA
I526A	**25.9 ± 6.6**	95.1 ± 11.7	98.8 ± 2.0	NA
T527A	**68.3 ± 9.9**	89.9 ± 2.3	80.6 ± 9.6	NA
G528A	**27.3 ± 1.6**	88.7 ± 5.2	91.1 ± 4.1	NA
M529A	89.1 ± 10.3	105.6 ± 4.6	100.8 ± 2.0	NA
V530A	**61.0 ± 32.7**	102.7 ± 2.2	99.1 ± 2.9	NA
D531A	**35.0 ± 8.6**	90.3 ± 9.5	85.9 ± 3.9	NA
P532A	96.9 ± 15.3	105.5 ± 4.4	99.9 ± 1.6	NA
S533A	108.6 ± 7.0	108.2 ± 5.2	100.9 ± 2.6	NA
R534A	84.1 ± 18.2	104.5 ± 8.1	98.4 ± 2.6	NA
I535A	**40.1 ± 29.0**	86.8 ± 4.7	81.5 ± 7.4	NA
N536A	**72.3 ± 30.1**	102.0 ± 7.6	99.4 ± 1.0	NA
V537A	**28.4 ± 18.2**	**67.3 ± 16.6**	**29.0 ± 14.0**	**31.2±15.1**
N539A	100.7 ± 12.3	103.6 ± 8.2	97.6 ± 2.4	NA
L540A	92.7 ± 17.0	100.8 ± 5.2	96.6 ± 3.0	NA
E542A	96.5 ± 4.9	100.0 ± 8.9	96.4 ± 0.3	NA
E543A	**67.9 ± 19.6**	99.6 ± 10.1	99.8 ± 1.3	NA
G544A	76.4 ± 22.7	104.2 ± 2.4	100.2 ± 2.3	NA
L545A	**57.1 ± 13.0**	83.6 ± 10.3	79.7 ± 1.2	NA
H610A	**3.0±1.1**	**0.1±0.1**	**10.4±12.1**	**1.9±2.4**
C613A	**0.2±0.2**	**0.0±0.0**	**5.6±6.6**	**1.6±2.2**
D632A	**0.1±0.1**	**0.1±0.2**	**0.0±0.0**	**1.2±1.2**
C635A	**0.1±0.1**	**1.0±0.8**	**11.1±7.7**	**1.0±0.9**
P638A	**2.2±3.0**	**2.9±2.0**	**31.6±15.5**	**15.5±6.8**

37ºC, the great majority exhibited severe folding defects. As the temperature was lowered, many single alanine mutants were rescued back to wild-type-like folding efficiencies. However, even at the low temperatures of 18ºC, 36 mutants were folding deficient and were unable to reach the native state like wild-type.

Mutants classified as folding deficient at 18ºC were subsequently expressed at 10ºC. The overall expression efficiencies observed at 10ºC were consistent with those seen at 18ºC. The lowered expression temperature did not rescue any of these residues to wild-type-like folding efficiencies.

4. Wild-type and ΔN Folding Efficiencies

For wild-type chains expressed at 18ºC, 94% (average of 22 expression samples) of the polypeptide chains reached the SDS-resistant, trimeric native conformation. When renormalized to this average, the standard deviation of these samples was 2%. Expression at 10ºC generated a folding yield of 93±4% (6 samples). When expressed at 30ºC, 87±6% (21 samples) of wild-type chains reached the native conformation. Expression at 37ºC yielded 62±10% (22 samples) of wild-type chains reaching the native state. The deviations at each temperature were well within the 75% threshold for a wild-type-like folder (Figure 2-3 insets; Table 2-2). The small standard deviations observed at each expression temperature illustrate the precision of this SDS-PAGE based assay.

ΔN chains folded with wild-type-like efficiency at 37ºC, 95±13% for 28 samples. At lower temperatures, ΔN folded significantly better than wild-type with a folding efficiency relative to wild-type of 109±3% for 26 samples at 30ºC, 104±1% for 28 samples at 18ºC, and 101±4% for 6 samples at 10ºC. It has been reported that tailspike chains lacking the N-terminal head-binding domain fold faster than the full wild-type sequence (Danner et al., 1993; Miller et al., 1998b). The ability to detect the affects of this accelerated folding rate validates the precision and accuracy of this assay.

5. Mutant Folding Efficiencies at 37ºC

In contrast to the insignificant effect that removal of over 100 amino acids had on the folding of the ΔN tailspike protein, alanine substitution of any one of 95 of the 103 buried core sites significantly hindered the tailspike chains' ability to reach the native state when expressed at 37ºC. One of these mutants, F308A, was previously shown to fail to fold at high and low temperatures (Betts et al., 2004). Our high throughput system was able to accurately reproduce the published data for this residue as well as all other β-helix residues investigated by Betts, et al (2004) (Table 2-2). Of the remaining 42 predominantly surface exposed sites, 20 had defective folding phenotypes. When folding data for all mutants are mapped to the native structure of the P22 tailspike (Figure 2-4), it is apparent that the vast majority of defective folding sites are in the cylindrical buried core. These sites were intolerant to alanine substitution when folded at 37ºC *in vivo*.

Conversely, the 30 wild-type-like folding alanine mutations map almost entirely to surface exposed residues, to the termini of the β-helix, or to the shortened stack that precedes sheet C. V386A and V455A, located in a valine stack in sheet B, were the only exceptions to this. G511A, located in a disordered loop in the wild-type crystal structure, had a folding efficiency of 86±9% at 37ºC, and remained a wild-type-like folder at lower temperatures.

6. Mutant Folding Efficiencies at 30ºC

At 30ºC, alanine substitutions at 47 sites prevented folding, as compared to 95 sites at 37ºC. Thus, a drop of 7ºC in expression temperature rescued 48 defective folding mutants back to wild-type-like levels. Amongst these rescued sites are nearly all the remaining surface exposed residues. 33 of the rescued positions were in the buried core stacks, so that 62 of the 103 stacking residues were intolerant to alanine substitution at 30ºC. Intolerant sites localized to the central region of the β-helix opposite the dorsal fin (Figure 2-4), while more tolerant positions clustered near the termini of the β-helix. The fourth and fifth rungs of the β-helix were completely intolerant to alanine substitutions of core residues at 30ºC.

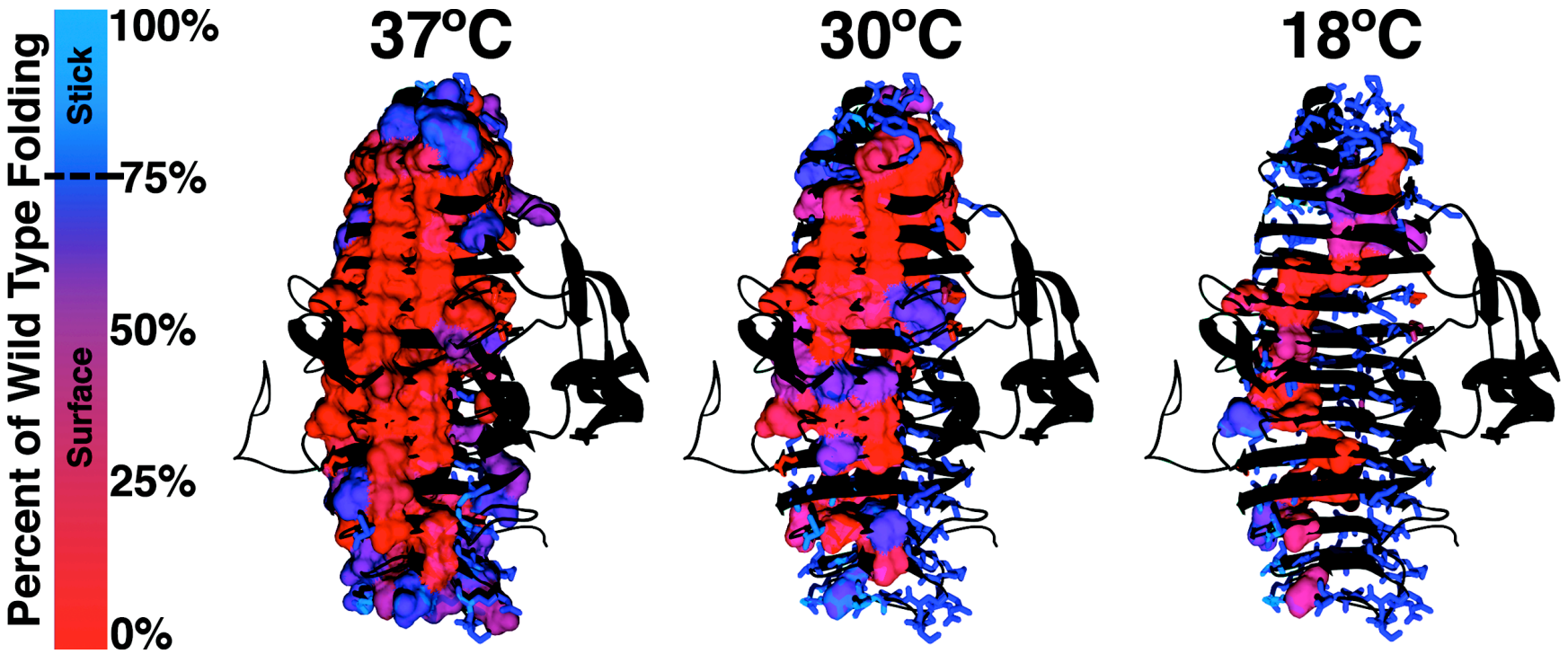

Figure 2-4. Folding efficiencies mapped onto the native structure of the tailspike β-helix. Mutated positions are colored according to the folding efficiency of the residue's alanine mutant. Positions that are glycine in the wild-type sequence or fold like wild-type are shown as stick representation. Folding deficient positions are represented as a surface volume. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

7. Mutant Folding Efficiencies at 18ºC

In contrast to the requirement for 92% of the buried core residues at 37ºC, only 31 out of 103 alanine mutations were absolutely incapable of achieving wild-type-like folding efficiencies at 18ºC. The sites of these completely intolerant residues were located in every rung of the β-helix, except rung 7. Rungs 6 and 8 have the highest number of intolerant positions, four and five respectively. Individual substitutions of six glycines prevented mutant chains from reaching the native state. Four of these six glycine residues have backbone conformations that are inaccessible by alanine, while steric packing problems may prevent the other glycine to alanine mutations from folding.

When non-glycine positions completely intolerant to alanine substitution are mapped onto the native structure of tailspike's β-helix, a continuous network of 29 interacting side chains is revealed that spans the length of the β-helix (Figure 2-4). This chain of mostly hydrophobic residues, which we call the "folding spine", is composed of short stacks that cluster around variable turn regions, such as the dorsal fin. In the case of the dorsal fin domain, a cluster of six amino acids is essential for folding where the chain detours from the β-helix and into the dorsal fin. At the center of this cluster, two residues (M195, L262) pack against each other, as if stapling together the meandering backbone (Figure 2-5).

Although the most frequent wild-type residue found in the folding spine is phenylalanine, it is important to note that not all phenylalanines in the buried core of the P22 tailspike β-helix are necessary at all temperatures. This observation demonstrates that the folding phenotype is not due solely to the type of mutation, but that tertiary context contributes to the detected folding efficiencies.

Two asparagines, N310 and N319, and the only histidine in the buried core, H426, are present in the folding spine. In the native structure, these side chains interact with resolved, buried water molecules that coordinate the backbone of loop regions. These necessary residues highlight the complexity of the folding process *in vivo*.

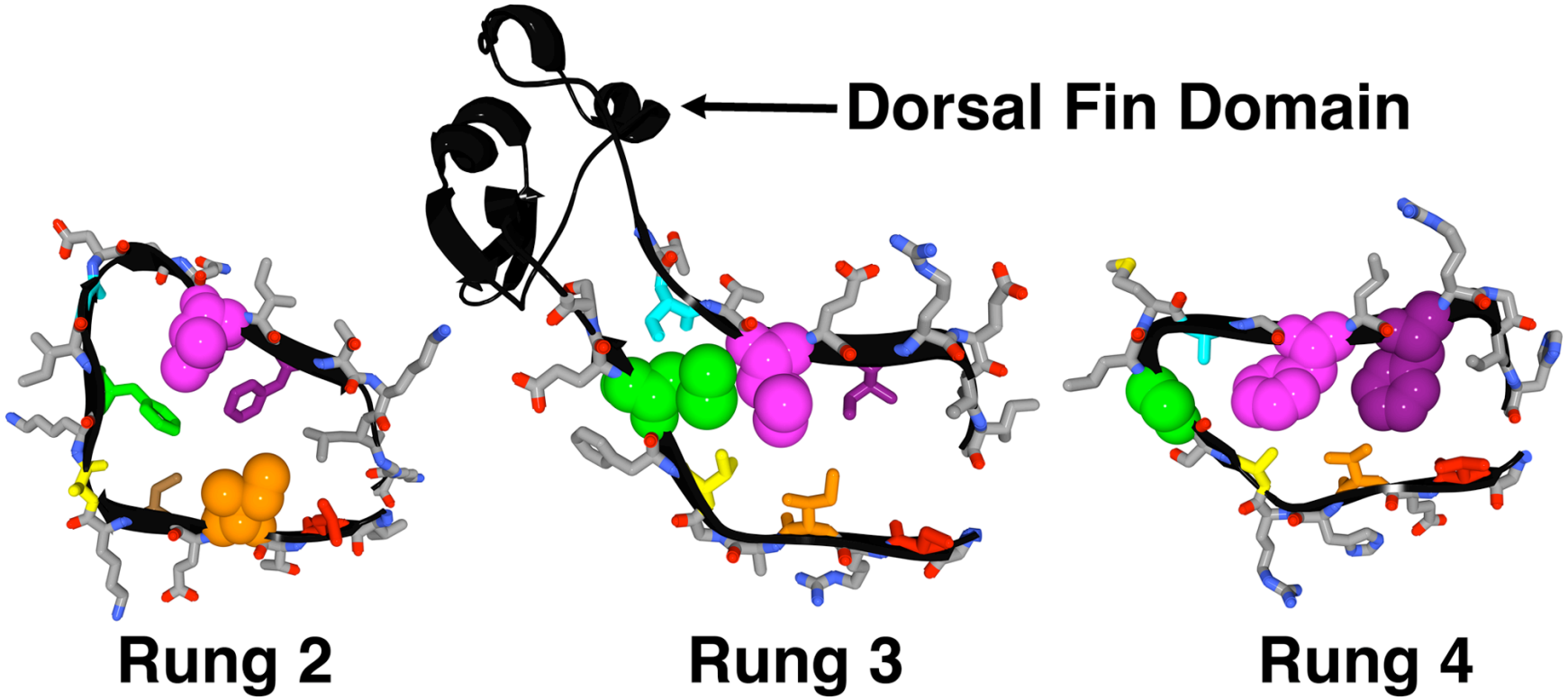

Figure 2-5. Critical residues surrounding the dorsal fin deviation from the β-helix. Residues are colored according to the stack color scheme in Figure 2-1. Residues critical for folding at 18ºC and 10ºC are shown as Van Der Waal's representation, while all other residues are represented by sticks. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

8. Mutant Folding Efficiencies at 10ºC

When expressed at 10ºC, the 31 buried core alanine mutations that were incapable of achieving wild-type-like folding efficiencies at 18ºC continued to display folding-deficient phenotypes (Table 2-2). Of these 31 residues, only 7 showed statistically different folding efficiencies, all but one of which had lower folding efficiencies at 10ºC than at 18ºC. The only mutant that significantly improved in its folding efficiency at 10ºC was L179A, which had its folding efficiency rise from 47% to 69% with the eight degree decrease in temperature.

9. Capping Residues

A number of alanine substitutions in the N-terminal capping region (122-146) of the β-helix resulted in folding deficient phenotypes at or below 30ºC. These mutants highlight this region's role in preventing aggregation (Figure 2-6). This region is composed of a short β-strand, a turn, and a conserved α-helix. V141A, located at the C-terminus of the conserved α-helix, is the only mutant in this region deficient in folding at all temperatures. This residue may act as a stop signal, ensuring that the α-helix is the proper length to cap the β-helix.

Only two capping mutants are rescued at 18ºC, L133A and Y130A. L133 is located at the N-terminus of the α-helix and points directly downward into the β-helix core, possibly anchoring the α-helix into its capping position. Y130 sits centrally over the core of the β-helix, potentially bridging the hydrophobic interior of the β-helix and the polar solvent. It is important to note that Y123A, as opposed to Y130A, displays completely wild-type-like folding phenotypes at all temperatures. The folding phenotypes observed in the N-terminal capping regions illuminate key roles of wild-type side chains that are not simply a reflection of the type of mutation.

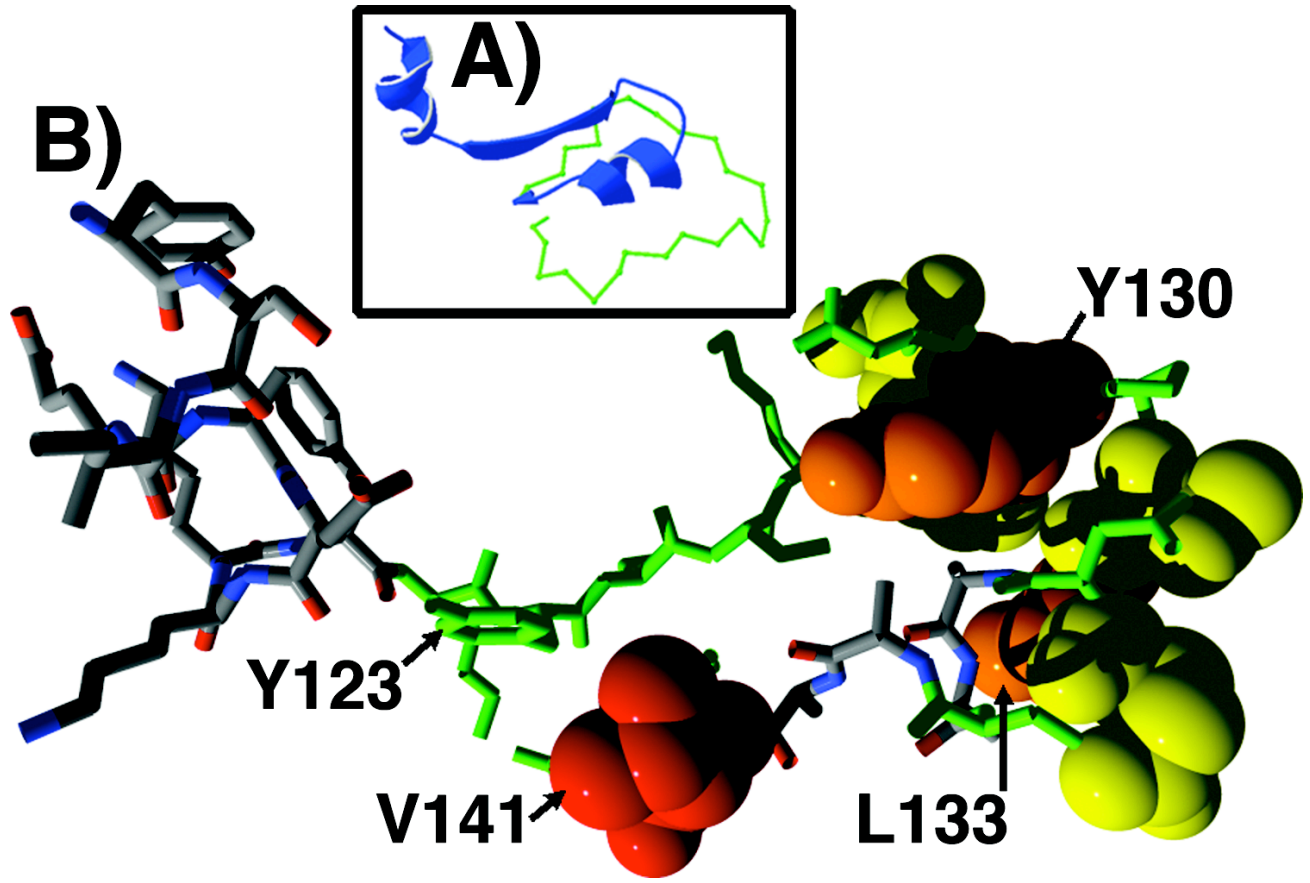

Figure 2-6. N-terminal Capping Region. (A) Ribbon diagram of the N-terminal capping region (blue) and α-Carbon trace of the first rung of the β-helix (green). (B) Phenotypic data is mapped onto the structure of the N-terminal capping domain. Green residues represent positions that fold like wild-type at all temperatures when replaced by alanine. Yellow residues are folding deficient only at 37ºC when replaced by alanine. Orange residues are folding deficient at 37ºC and 30ºC when replaced by alanine. The red residue (V141) is folding deficient at all temperatures when replaced by alanine. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

In the C-terminal capping region (521 – 545), only V537A was deficient in folding at temperatures below 37ºC (Table 2-2). In the wild-type native trimer, this residue is a part of the final rung of the β-helix and interacts with the side chain of F553 from one of its sister polypeptide chains. This residue may play a greater role in capping through oligomerization, rather than capping in a monomeric state.

10. C-terminal Cysteine Annulus Mutations

Alanine substitutions were constructed in the far C-terminus of the tailspike protein. These mutants are located in an area of the structure known as the cysteine annulus, which is known to be important for the oligomerization of P22 tailspike (Gage and Robinson, 2003). These mutations affected cysteine residues (C613 and C635) believed to be important for the presence of a transient disulfide bond in a protein folding intermediate (Haase-Pettingell et al., 2001; Robinson and King, 1997) as well as residues that appeared likely to affect the reactivity of those cysteine residues (H610, D632, and P638) (Fomenko and Gladyshev, 2003).

At all temperatures tested, these mutants were incapable of folding to the native, trimeric state (Table 2-2).

11. Solubility and Mobility Shifts

Instead of accumulating as trimers, the majority of the cysteine annulus mutations accumulated as SDS-sensitive polypeptide chains in the soluble fraction of the expression lysates at temperatures below 37ºC. At 37ºC, all non-native tailspike conformers in all 37ºC samples accumulated in a consistent manner in the pelleted, insoluble fraction of the cell lysates. At 30ºC and lower temperatures, only V537A, H610A, C613A, and D632A accumulated soluble, SDS-sensitive tailspike chains in a consistent manner across the three independent expression cultures tested per mutant. C635A also accumulated soluble, SDS-sensitive chains at 18ºC and 10ºC in a consistent manner.

As described above, accumulation of soluble SDS-sensitive species represents chains that have reached a late stage in folding, with defects either in oligomerization,

native state maturation, or stability. Since mutants that accumulated as soluble, SDS-sensitive polypeptide chains are located at the very end of the β-helix or in the C-terminus of the tailspike protein, it is consistent to state that these chains were capable of creating a folded β-helix domain but were defective in later steps in the tailspike folding pathway. In contrast, the folding defects associated with the core β-helix alanine mutants occurred at or prior to the monomeric, β-helical folding intermediate.

A number of native trimeric mutant chains displayed mobility shifts with respect to wild-type chains. These mutants primarily corresponded to the mutation of charged residues, though not all charge changes produced alterations in mobility. Two neutral changes (N499A, G462A) produced reproducible mobility shifts. The physical basis for these shifts is not understood, but may reflect an altered native conformation.

D. DISCUSSION

1. Tailspike β-helix Folds via a Processive Folding Pathway

Previous experiments on small, globular proteins have demonstrated that the amino acid folding code is highly degenerate and, in at least one case, half of the residues can be replaced by alanines simultaneously (He et al., 2004; Kuroda and Kim, 2000; Sauer, 1996). In the P22 tailspike β-helix, the entire core is surprisingly sensitive to amino acid changes at physiological temperature. If folding followed a global collapse mechanism, then the removal of 2 to 8 atoms out of a core composed of 365 side chain atoms would not be expected to completely block the folding process, which has a ΔG of -7.6 kcal/mol for the isolated β-helix domain (Miller et al., 1998a) and produces a native state that has a melting temperature of 88.4ºC (Sturtevant et al., 1989).

In contrast, a processive folding pathway, where a series of local folding events must occur in a specific order to reach a native structure, is consistent with these results. We propose that the *in vivo* folding pathway of the tailspike's β-helix occurs via a processive folding mechanism, such as the sequential winding of rungs (Figure 2-7).

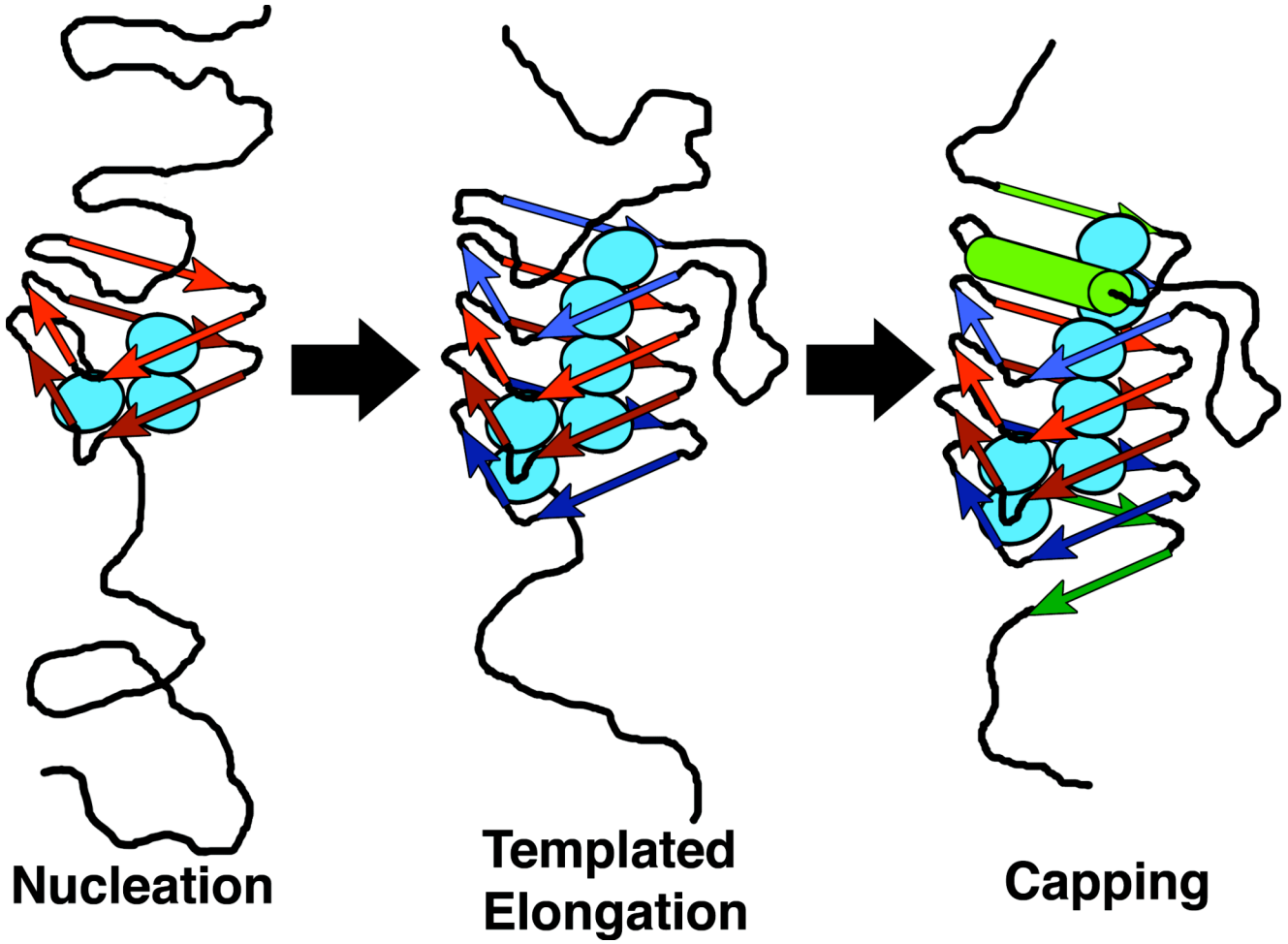

Figure 2-7. Model of the processive folding of the β-helix. Critical internal side chain contacts (cyan ovals) nucleate an initial set of well-formed rungs. This nucleus acts as a template for other rungs to assemble via side chain stacking and/or tethering around loop regions. Capping of the N-terminus likely prevents aggregation. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

The contribution of wild-type side chains to productive folding was temperature dependent. Folding analysis at low temperatures revealed an elongated folding spine of essential residues encoding critical contacts that allow for the *in vivo* folding of the tailspike protein. Although it cannot be ruled out that these residues represent positions that critically interact with the ribosome in a ribosome-assisted folding mechanism, it is unlikely that these residues are required for intermolecular interactions. Instead, we postulate the simpler model that these residues may control the folding of the β-helix by tethering key positions into a general cylindrical structure, thereby allowing β-sheets to form. At lower temperatures, certain packing interactions may tolerate an alanine replacement, but the interaction energy of an alanine mutant may not be sufficient to overcome added kinetic energies at higher temperatures.

30ºC data demonstrated that residues located centrally in the β-helix were more intolerant to substitution than at the termini. The proposed processive folding pathway predicts that inhibition of earlier folding events would have more drastic effects than later stages. This suggests that β-helix formation *in vivo* initiates with the creation of a structural nucleus near the center of the β-helix and proceeds via a templating or tethering mechanism outward towards the termini. Folding data from mutants located in the N-terminal capping region suggest that folding of this domain concludes via formation of key capping contacts, which may stabilize the β-helix and prevent aggregation.

All β-helix domain mutant chains incapable of producing native trimers accumulated in the insoluble pellet fraction of the lysates, while mutations in the C-terminus of the protein accumulated soluble, SDS-sensitive chains. This indicates that the observed folding defects associated with mutants in the β-helix domain are due to affects on the folding of this domain.

2. Role of Stacking Residues

These results indicate that stacking of side chains plays a critical role in the folding of the β-helix. However, stacks that run the entire length of the β-helix are not necessary to direct folding at low temperatures. Instead, the folding spine spans numerous stacks and is composed of numerous different amino acid types. This finding supports the notion that side chain stacks are necessary to encode an elongated β-helix. This is in contrast to the large number of previously isolated *tsf* mutations that represent solvent exposed residues critical in folding at high temperatures, but unnecessary at lower temperatures (Betts et al., 2004). The short core stacks may function to stabilize intermediate forms or to specifically prevent off-pathway aggregation. Although the identity of wild-type side chains in the folding spine does not yield a simple sequence-based rule for folding, it does reveal that folding is a highly complex process that depends both on side chain identities and their tertiary context.

The continuous nature of the folding spine indicates that a long-range network of interactions may act cooperatively to form the β-helix. Such networks have been previously proposed based upon sequence conservation analysis of PDZ and POZ domains (Lockless and Ranganathan, 1999). Our data demonstrates the existence of such a network in the context of the folding of the β-helix domain. If the connectivity of short stacks did not act cooperatively, one might expect that these short stacks of necessary residues might be more randomly dispersed throughout the core. Instead, we believe that the presence of this continuous folding spine represents a cooperative network that directs the protein towards its native fold *in vivo*.

Understanding the full role key stacking positions play will require further investigations into the cooperativity of this network and identification of factors that link the amino acid sequence to its contribution towards folding. Seckler and coworkers have analyzed the site of the buried global suppressor, A334, in considerable detail (Seckler, 1998). In contrast to the intolerance of side chain deletions presented here, β-helix folding tolerated bulkier substitutions at this site (Schuler and Seckler, 1998).

3. Capping Residues

Residues located at the N-terminus of the β-helix indicate that this capping region is critical for prevention of aggregation. V141 likely acts as an α-helix breaker, thereby ensuring the formation of the capping α-helix. The positioning of Y130 over the core the β-helix not only acts as a bridge between the hydrophobic interior of buried core and the polar solvent, but it may also register the α-helix into its capping role on the N-terminal side of the helix. Overall, it appears that the formation of the α-helix and the positions bounding it are critical for the proper formation and solubility of the β-helix.

In contrast, residues located at the C-terminus of the β-helix did not appear to be as important for the folding of the protein. If these residues can tolerate alanine substitutions, how does the sequence encode a signal to stop the winding of the β-helix domain and to protect the hydrophobic surface that may be exposed until the protein trimerizes? The answer may lie in the fact that the region of sequence following the last rung of the β-helix contains a large number of charged residues. These residues are unlikely to be able to form a rung without causing burial of charged residues (Figure 2-8). Packing of buried charges would be incompatible with the sole critical residue in this region, V537. Therefore, the C-terminal residues that were tested in this experiment were likely unimportant for folding because the true "stop" signal exists beyond the sequence region examined.

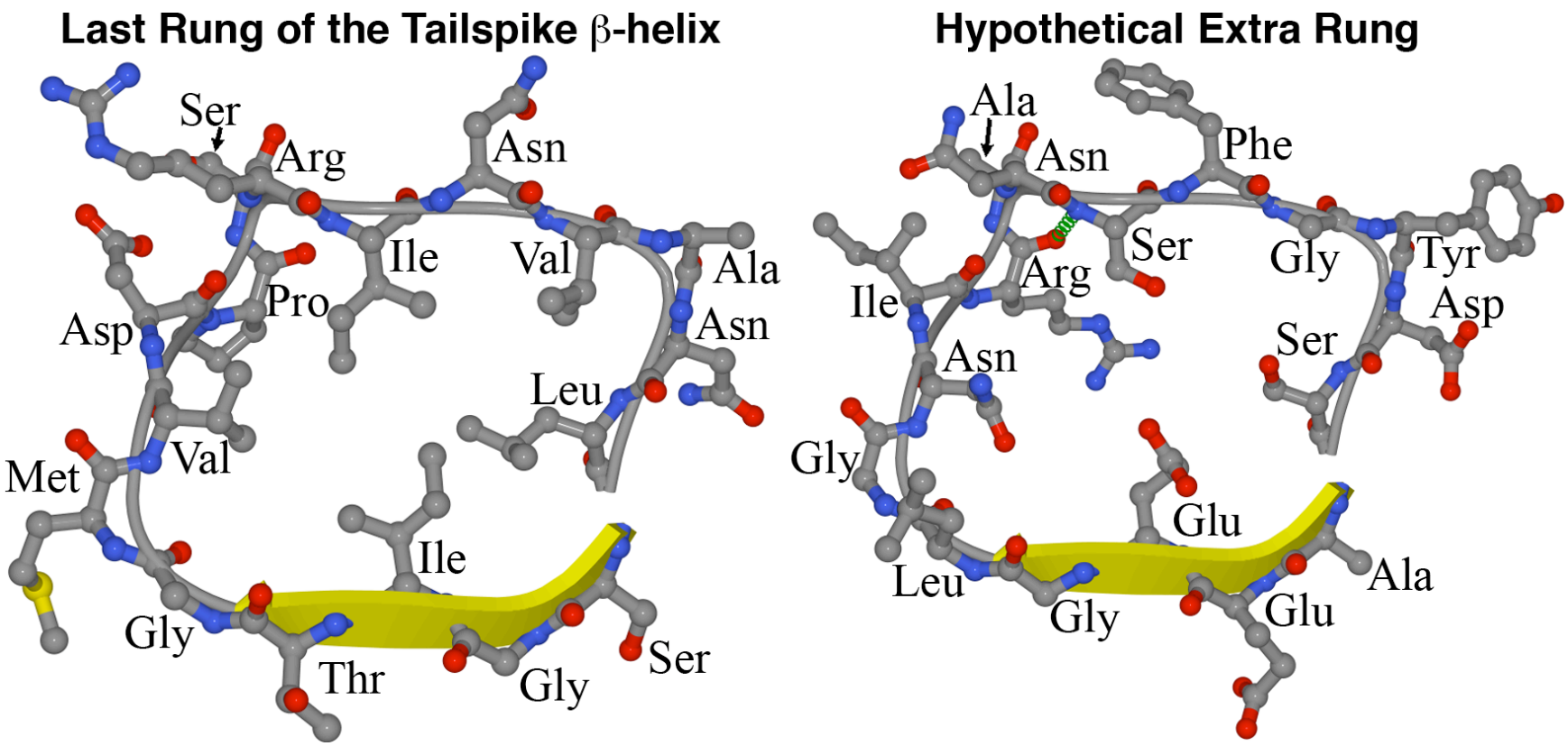

Figure 2-8. C-terminal Capping Region. The left image depicts the structure of the last rung of the tailspike β-helix, from residue 524 to 540, as observed in the PDB file 1TYU (Steinbacher et al., 1996). The right image depicts a threaded model of residues 541 to 557 onto the backbone seen on the left. These residues normally form the triple β-helical, interdigitated region of the tailspike trimer.

CHAPTER THREE:

THERMOSTABILITY OF FOLDING MUTANTS

A. INTRODUCTION

1. Thermostability of P22 Tailspike

The last step in tailspike's productive folding pathway (Figure 1-8) coincides with the folding and assembly of the interdigitated region to form a molecular clamp that makes tailspike resistant to trypsin degradation and SDS denaturation, and raises the melting temperature by approximately 40°C to 88.4°C (Goldenberg and King, 1982; Kreisberg et al., 2002; Sturtevant et al., 1989). When exposed to temperatures above 65ºC and below tailspike's melting temperature in the presence of 2% SDS, the native trimer will partially unfold to an SDS-resistant, high mobility species (I_u) that maintains β-helical conformation but lacks structure in the N-terminal head-binding domain (Chen and King, 1991).

Previous experimentation on temperature sensitive folding mutations has demonstrated that temperature sensitive folding (*tsf*) mutations do not affect the native stability of the trimeric tailspike protein, but instead affect a thermolabile folding intermediate (Danner and Seckler, 1993; Sturtevant et al., 1989).

2. Experimental Design

In Chapter 2, it was shown that polypeptide chains carrying a β-helix mutant incapable of producing a native trimer band accumulated in the pelleted fraction of the lysate. In contrast, folding deficient mutants located at the very end of the β-helix or in the cysteine annulus of the tailspike were capable of producing soluble but SDS-sensitive tailspike chains. Although this indicates that β-helix mutants affected the folding of the protein before association, it does not rule out the possibility that mutant polypeptide chains reached a transient native-like state, but then fell apart to an aggregation prone state. Do these mutations destabilize the native trimeric conformation of tailspike?

To test if these mutations caused low folding efficiencies at the higher 37ºC temperature by destabilizing the native state, I performed a thermostability assay on mutant trimers produced at 18ºC to determine their approximate melting temperatures.

B. MATERIALS AND METHODS

1. Production of Mutant Trimers

Natively folded tailspike trimers of mutants generated in Chapter 2 were produced by over expression at 18ºC as described in Chapter 2, Section B2. Cells were lysed and fractionated by centrifugation as described in Chapter 2, Section B3. The supernantant fractions containing native trimers were subsequently mixed with 2% SDS sample buffer as described in Chapter 2, Section B3 in a 96 well plate. After the experiments of Chapter 2 were completed, these samples were stored at -20ºC.

2. Thermostability Assay

Frozen supernatant samples were thawed on ice and 10 µl aliquots were moved to new shallow well plates. Plates were sealed with silicon cap lids clamped down by a glass plate and binder clips. The sealed plate was partially submerged in a heated water bath to either 80ºC or 90ºC for 3 min, and then placed on ice until samples were examined by SDS-PAGE, as described in Chapter 2, Section B3. Gels were imaged and analyzed by optical densitometry as described in Chapter 2, Section B4. Due to the partial unfolding of the N-terminal domain of tailspike at temperatures above 65ºC (Chen and King, 1991), optical density values were calculated for the tailspike bands corresponding to the native trimer, the partially unfolded intermediate I_u, and the completely unfolded SDS-sensitive species. The percentage of chains that were unfolded was calculated by dividing the optical density of the completely unfolded band by the total optical density of all three tailspike bands, as detailed in the following formula:

$$\text{Percent Unfolded} = \frac{OD_U}{OD_{NT} + OD_{I_U} + OD_U} * 100\%$$

C. RESULTS

1. Unfolding of Wild-type and ΔN Controls

To test if mutations acted by destabilizing the native state, the native trimeric forms of the mutant polypeptide chains in the soluble fractions of 18ºC expression samples were assessed for thermostability. 114 of the alanine mutants accumulated sufficient native-like trimers at low temperature to be assayed. Lysates were heated to either 80ºC or 90ºC for 3 min in the presence of 2% SDS and analyzed for unfolding of the β-helix region. As described by Chen and King (Chen and King, 1991), wild-type proteins partially unfolded to an SDS-resistant, high mobility species (I_u) that maintains β-helical conformation but lacks structure in the N-terminal head-binding domain when heated to 80ºC (Figure 3-1). Only 3.1±1.7% of wild-type native chains fully denatured to a SDS-sensitive species at 80ºC. Heating to 90ºC resulted in nearly complete dissociation (89.9±7.9%) to a denatured, monomeric SDS/polypeptide chain complex (Figure 3-2; Table 3-1).

Heating of ΔN resulted in similar thermostability values, with only 4.5±3.4% of chains unfolding at 80ºC and 98.5±3.7% of chains unfolding at 90ºC (Figure 3-2; Table 3-1). Since ΔN lacks the N-terminal region that partially unfolds at these high temperatures, ΔN did not produce the partially unfolded species (I_u) (Figure 3-1). This reconfirmed that the partial unfolding of the mutant trimers was not due to the unfolding of the β-helix region and allowed for the identification of mutant tailspike specific bands, such as I_u, in other samples.

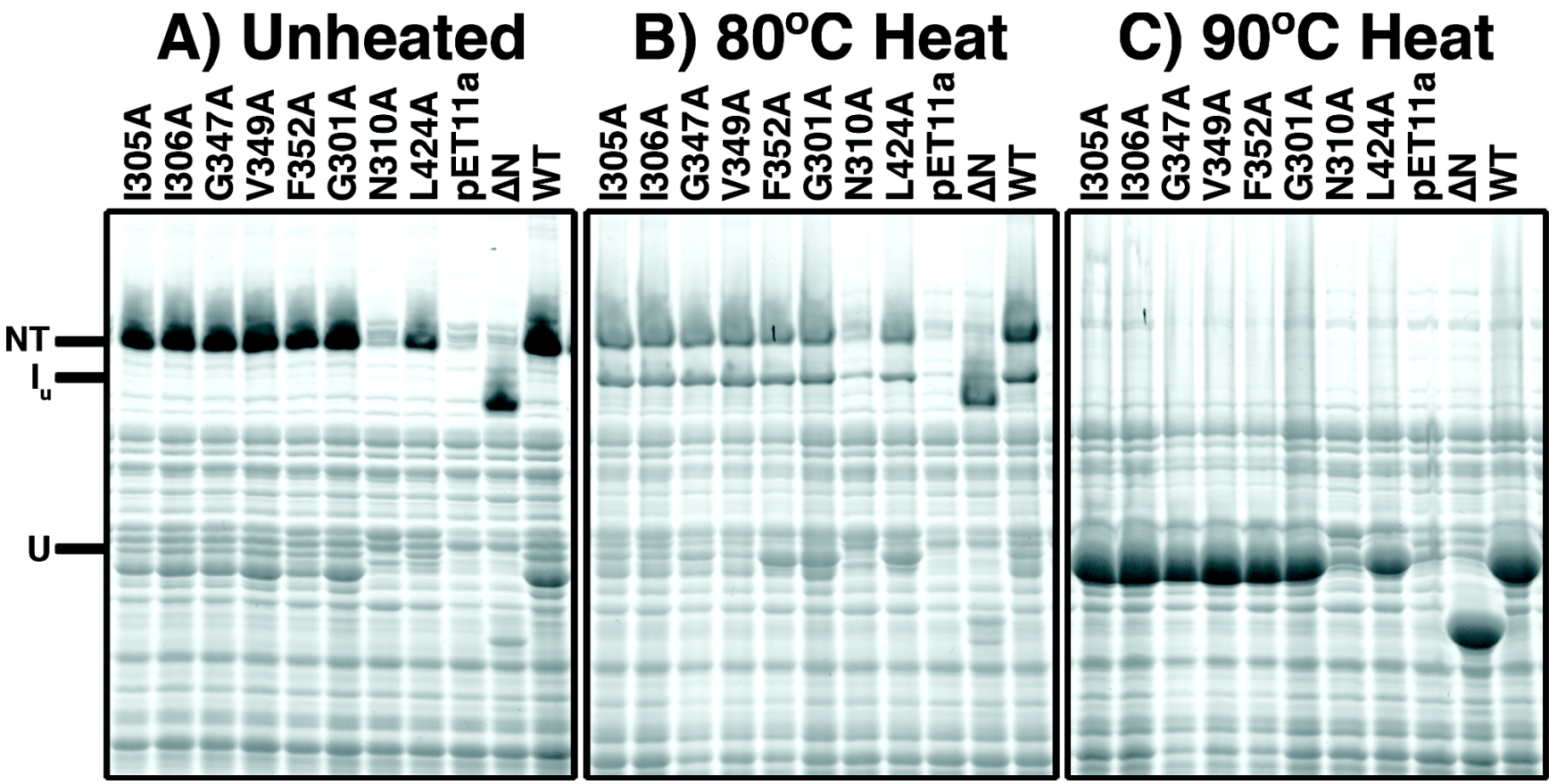

Figure 3-1. Sample SDS-PAGE data from thermostability assays. (A) Unheated supernatant samples from 18ºC expressions. (B) Samples after heating to 80ºC. (C) Samples after heating to 90ºC. NT = SDS-resistant, trimeric native tailspikes; I_U = Trimeric tailspikes with partially unfolded N-terminal head binding domains; U = Tailspikes that have been dissociated and unfolded in the presence of SDS. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

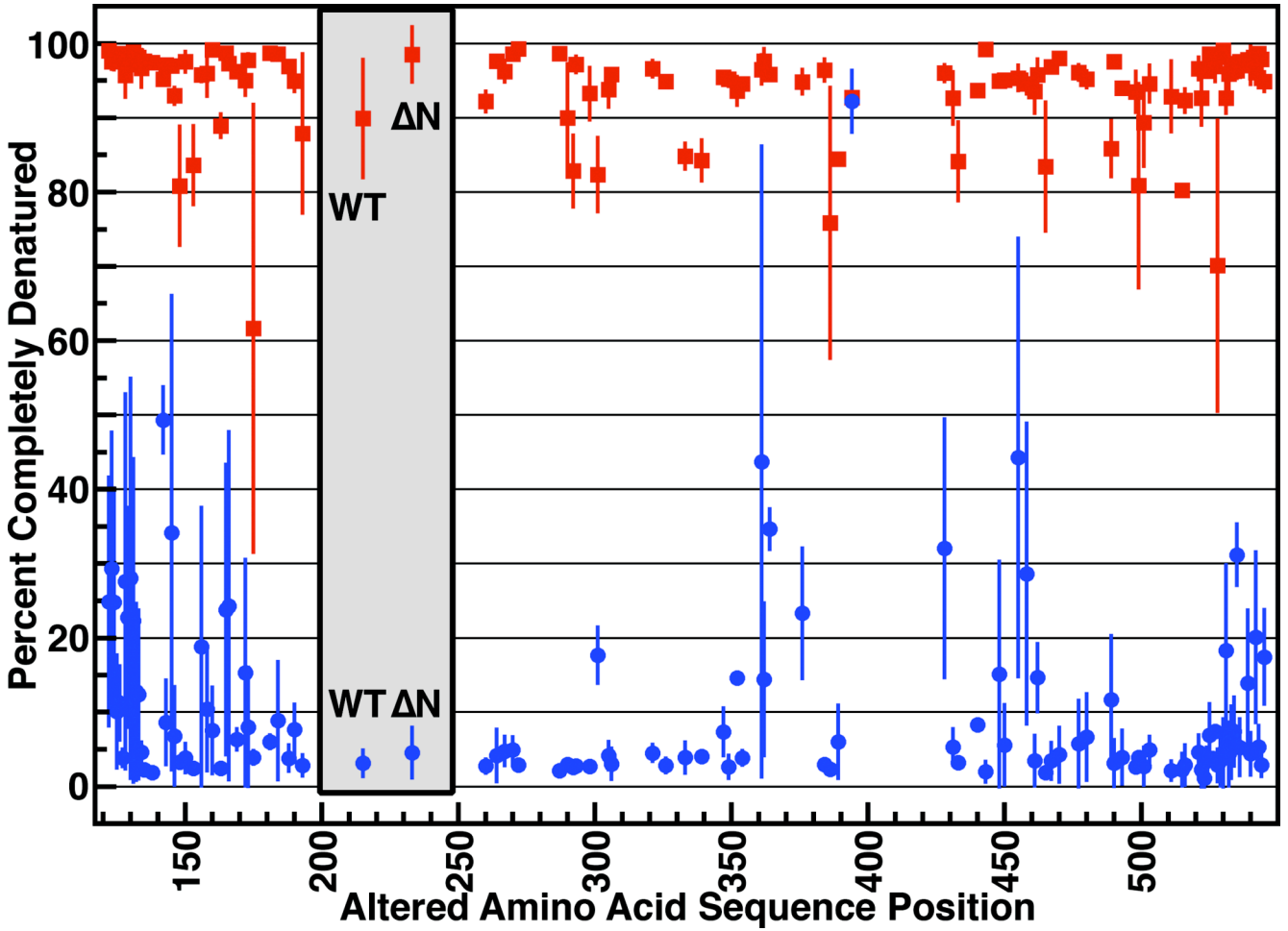

Figure 3-2. Thermostability assay on 18ºC wild-type-like folders. The percent of native tailspike chains completely unfolded in the presence of 2% SDS when heated to 80ºC (blue circle) or 90ºC (red square) is shown as the average of duplicate experiments. Lines represent standard deviations. WT and ΔN controls are shown in inset. Figure reprinted from (Simkovsky and King, 2006), © 2006 by The National Academy of Sciences of the USA.

Table 3-1. Thermostability data of 18ºC WT-like trimers. Thermostability values are given as the percent of tailspike chains completely unfolded in the thermostability assay when heated to either 80ºC or 90ºC. These values were calculated only for samples displaying >75% folding efficiency at 18ºC, as determined in Chapter 2. Otherwise the calculation was not applicable (NA). Average percent unfolded values greater than 50% are shaded and bolded.

Sample	Percent unfolded by heat	
	80ºC heat	90ºC heat
WT	3.1 ± 1.7	**89.9 ± 7.9**
ΔN	4.5 ± 3.4	**98.5 ± 3.7**
K122A	24.9 ± 16.7	**99.0 ± 0.6**
Y123A	29.3 ± 18.3	**97.6 ± 1.0**
S124A	24.8 ± 15.4	**97.8 ± 1.3**
V125A	10.1 ± 7.6	**98.2 ± 0.5**
K126A	11.2 ± 5.0	**97.9 ± 0.0**
L127A	3.9 ± 1.1	**98.6 ± 0.0**
S128A	27.6 ± 25.2	**95.7 ± 2.9**
D129A	22.8 ± 14.8	**97.2 ± 0.2**
Y130A	28.0 ± 26.9	**97.4 ± 0.4**
P131A	22.4 ± 21.7	**98.8 ± 0.1**
T132A	12.8 ± 11.9	**98.3 ± 1.0**
L133A	12.4 ± 11.3	**98.2 ± 0.7**
Q134A	4.6 ± 1.3	**96.7 ± 2.5**
D135A	2.3 ± 0.3	**97.6 ± 0.9**
S138A	1.9 ± 0.2	**97.4 ± 0.8**
V141A	NA	NA
D142A	49.3 ± 4.4	**95.2 ± 0.0**
G143A	8.6 ± 5.7	**97.1 ± 0.2**
L144A	NA	NA
L145A	34.2 ± 31.9	**97.0 ± 0.3**
I146A	6.8 ± 6.6	**93.0 ± 1.1**

Sample	Percent unfolded by heat	
	80ºC heat	90ºC heat
R148A	3.3 ± 0.4	**80.8 ± 8.0**
Y150A	3.8 ± 2.0	**97.6 ± 1.4**
F152A	NA	NA
Y153A	2.5 ± 0.3	**83.6 ± 5.3**
E156A	18.8 ± 18.7	**95.8 ± 0.9**
V158A	10.4 ± 8.3	**96.0 ± 3.1**
F160A	7.5 ± 5.8	**99.1 ± 0.1**
K163A	2.5 ± 0.2	**88.9 ± 1.5**
L165A	23.8 ± 19.5	**98.7 ± 0.4**
T166A	24.3 ± 23.4	**97.3 ± 0.7**
I167A	NA	NA
C169A	6.4 ± 1.4	**96.2 ± 0.1**
K172A	15.3 ± 15.2	**95.0 ± 2.0**
F173A	8.0 ± 8.0	**97.8 ± 0.8**
G175A	3.9 ± 0.9	**61.6 ± 30.1**
L179A	NA	NA
F181A	6.0 ± 0.9	**98.7 ± 0.5**
L184A	8.9 ± 7.9	**98.6 ± 0.5**
S188A	3.8 ± 1.8	**96.9 ± 0.6**
I190A	7.7 ± 3.4	**95.0 ± 1.4**
V193A	2.9 ± 1.4	**87.9 ± 10.7**
M195A	NA	NA
S260A	2.8 ± 0.9	**92.2 ± 1.3**

Sample	80ºC heat	90ºC heat
L262A	NA	NA
I264A	4.2 ± 3.5	**97.7 ± 0.6**
C267A	4.8 ± 1.9	**96.2 ± 1.4**
V270A	4.9 ± 1.7	**98.6 ± 0.0**
V272A	2.9 ± 0.7	**99.3 ± 0.3**
G277A	NA	NA
F282A	NA	NA
F284A	NA	NA
C287A	2.1 ± 0.5	**98.7 ± 0.1**
C290A	3.0 ± 0.5	**90.0 ± 7.3**
M292A	2.6 ± 0.6	**82.8 ± 4.8**
V293A	2.8 ± 0.7	**97.2 ± 1.1**
P298A	2.7 ± 0.1	**93.3 ± 3.5**
G300A	NA	NA
G301A	17.7 ± 3.8	**82.4 ± 5.0**
I305A	4.2 ± 1.9	**93.8 ± 2.4**
I306A	3.1 ± 2.0	**95.9 ± 0.5**
F308A	NA	NA
N310A	NA	NA
N319A	NA	NA
V321A	4.5 ± 1.1	**96.6 ± 1.1**
G324A	NA	NA
T326A	2.8 ± 0.9	**94.9 ± 0.7**
G329A	NA	NA
S333A	3.9 ± 2.1	**84.8 ± 1.7**
F336A	NA	NA
N339A	4.1 ± 0.5	**84.3 ± 2.8**
G347A	7.4 ± 3.2	**95.5 ± 0.1**
V349A	2.7 ± 1.5	**95.2 ± 1.0**

Sample	80ºC heat	90ºC heat
F352A	14.6 ± 0.5	**93.6 ± 1.9**
S354A	3.9 ± 1.0	**94.6 ± 0.3**
G361A	43.7 ± 42.4	**96.5 ± 2.0**
V362A	14.4 ± 10.3	**97.7 ± 1.6**
T364A	34.6 ± 2.7	**95.8 ± 0.3**
N376A	23.3 ± 8.7	**94.9 ± 1.6**
L379A	NA	NA
F381A	NA	NA
S384A	3.0 ± 0.6	**96.4 ± 1.5**
V386A	2.3 ± 0.7	**75.9 ± 18.2**
P389A	6.0 ± 4.9	**84.5 ± 0.3**
G393A	NA	NA
F394A	**92.2 ± 4.1**	**92.7 ± 2.2**
L396A	NA	NA
L424A	NA	NA
H426A	NA	NA
I428A	32.1 ± 17.4	**96.0 ± 1.1**
L431A	5.3 ± 2.4	**92.7 ± 3.5**
V433A	3.2 ± 0.2	**84.2 ± 5.3**
G440A	8.3 ± 0.2	**93.7 ± 0.7**
F441A	NA	NA
M443A	2.0 ± 1.4	**99.2 ± 0.1**
G445A	NA	NA
M448A	15.1 ± 15.1	**94.9 ± 0.6**
V450A	5.6 ± 5.4	**95.1 ± 0.6**
I453A	NA	NA
V455A	44.3 ± 29.5	**95.3 ± 1.7**
C458A	28.6 ± 20.2	**94.6 ± 1.3**
S461A	3.5 ± 3.4	**93.6 ± 2.9**

Sample	80ºC heat	90ºC heat
G462A	14.7 ± 4.6	**95.8 ± 2.1**
L465A	1.9 ± 0.3	**83.5 ± 8.6**
T467A	3.5 ± 2.4	**96.9 ± 0.7**
S470A	4.3 ± 3.6	**98.0 ± 0.7**
F472A	NA	NA
I475A	NA	NA
I477A	5.8 ± 5.8	**96.1 ± 1.0**
T480A	6.7 ± 5.8	**95.2 ± 1.1**
Q489A	11.7 ± 8.6	**85.9 ± 3.7**
I490A	3.2 ± 3.1	**97.6 ± 0.3**
I492A	NA	NA
S493A	3.9 ± 3.6	**94.0 ± 0.4**
C496A	NA	NA
V498A	2.7 ± 0.2	**93.5 ± 2.7**
N499A	3.9 ± 0.0	**80.9 ± 13.7**
L501A	2.7 ± 2.6	**89.4 ± 5.9**
L503A	5.0 ± 1.8	**94.6 ± 2.4**
G511A	2.1 ± 1.3	**92.9 ± 4.7**
T515A	2.3 ± 2.2	**80.2 ± 0.6**
I516A	2.9 ± 2.7	**92.3 ± 1.5**
S521A	4.6 ± 2.3	**96.6 ± 1.5**
T522A	2.3 ± 2.3	**92.7 ± 3.6**
V523A	1.1 ± 1.1	**96.4 ± 0.4**
S524A	3.4 ± 1.9	**96.3 ± 0.4**

Sample	80ºC heat	90ºC heat
G525A	6.9 ± 4.2	**98.6 ± 0.1**
I526A	3.8 ± 0.7	**97.6 ± 0.9**
T527A	7.4 ± 0.4	**96.9 ± 2.0**
G528A	2.9 ± 2.2	**70.1 ± 19.6**
M529A	3.5 ± 3.4	**98.2 ± 0.9**
V530A	4.6 ± 4.6	**99.1 ± 0.2**
D531A	18.3 ± 11.5	**92.7 ± 2.0**
P532A	4.4 ± 4.3	**95.9 ± 2.9**
S533A	5.9 ± 4.9	**96.8 ± 1.2**
R534A	7.4 ± 4.6	**96.3 ± 0.3**
I535A	31.2 ± 4.1	**96.3 ± 1.0**
N536A	5.3 ± 3.8	**97.6 ± 0.1**
V537A	NA	NA
N539A	13.9 ± 9.8	**97.8 ± 1.0**
L540A	4.4 ± 2.8	**97.1 ± 2.6**
E542A	20.1 ± 11.5	**96.1 ± 1.6**
E543A	5.3 ± 2.8	**98.7 ± 0.4**
G544A	2.9 ± 1.5	**97.9 ± 0.7**
L545A	17.5 ± 6.3	**94.9 ± 1.3**
H610A	NA	NA
C613A	NA	NA
D632A	NA	NA
C635A	NA	NA
P638A	NA	NA

2. Unfolding of Single Alanine Mutants

All trimeric mutant tailspikes consistently dissociated to monomeric SDS/polypeptide chain complexes when heated to 90ºC (Figure 3-2). All of these chains partially denatured similarly to wild-type when subjected to 80ºC, except for F394A. F394A consistently and completely dissociated to SDS-sensitive chains when subjected to 80ºC (Figure 3-2; Table 3-1). These results demonstrate that the single alanine mutants have melting temperatures within the range of 80ºC to 90ºC in the presence of 2% SDS. Therefore, all but one of the alanine mutations did not affect the stability of the trimeric native state.

D. DISCUSSION

1. Alanine Mutants Affect β-helix Folding Intermediates

In these thermostability experiments, mutant chains were exposed to extreme environmental conditions. The chains, while being present in a 2% SDS solution, were first frozen at -20ºC and then heated to either 80ºC or 90ºC. These settings are far harsher than the environment experienced by mutant chains folding within cells at 37ºC. Given that the melting temperatures of these chains is greater than 80ºC in the presence of detergent, it is clear that the observed folding efficiencies in Chapter 2 are not due to a destabilization of the native state, but must be due to a folding defect earlier in the productive folding pathway.

CHAPTER FOUR:

IN VIVO FOLDING OF β-HELIX MUTANTS IN THE ISOLATED β-HELIX DOMAIN

A. INTRODUCTION

1. The Isolated β-helix Domain

Seckler and coworkers constructed and characterized a tailspike fragment consisting of only the central β-helix domain (109 – 544) of the P22 tailspike protein (Miller et al., 1998a). They demonstrated that the fluorescence and CD signals of the isolated β-helix construct closely resemble those of the full length tailspike, indicating that the fragment is in a conformation similar to that of the β-helix domain in the native trimeric tailspike. This tailspike fragment exists as a monomer at concentrations below 1 μM, trimerizes at higher protein concentrations, and is capable of binding and cleaving tailspike's endorhamnosidase substrate. Unlike the full length tailspike, the β-helix fragment can be reversibly unfolded and refolded under ideal conditions, allowing for the calculation of its free energy of folding (-32 kJ/mol) from a two-state model.

Under non-ideal conditions, such as temperatures above 10ºC, high salt concentrations, or the lack of chelating agents, the β-helix fragment readily aggregates (Schuler et al., 1999). This aggregation proceeds by a linear polymerization model and occurs fastest at denaturant concentrations corresponding to its folding mid-point, indicating that aggregation proceeds from a partially folded intermediate. Aggregation leads to a number of distinct morphologies, including the formation of amyloid-like fibrils. These fibers bind the dye Congo red and display bright green birefringence upon excitation with cross-polarized light. Unlike most amyloids, the fibers are an order of magnitude larger in diameter and many of the fibers appeared curled or branched.

2. Mutants G244R *(tsf)* and A334V *(su)*

Mutational analysis of the full length tailspike and the isolated β-helix domain have allowed for detailed investigations into the sequence control of the folding of the β-helix. Two particular mutations, G244R and A334V, have been well characterized in both constructs under various environmental conditions *in vivo* and *in vitro*.

The temperature sensitive mutation (*tsf*) G244R, originally designated *tsH304*, is located in a solvent exposed loop of the dorsal fin. This mutation produces trimeric tailspike at temperatures at or below 30ºC with an altered electrophoretic mobility both *in vivo* and *in vitro* (Goldenberg and King, 1982; Goldenberg et al., 1982; Mitraki et al., 1993; Yu and King, 1984). At elevated temperatures, such as 37ºC or greater, G244R is incapable of forming trimeric tailspikes, even though trimeric G244R tailspikes formed at permissive temperatures have a melting temperature of 87.5ºC (Goldenberg and King, 1981; Smith and King, 1981; Sturtevant et al., 1989). Instead, a destabililzed folding intermediate rapidly accumulates into an irreversible aggregate, as is apparent from *in vitro* refolding experiments (Danner and Seckler, 1993). When investigated in the context of the isolated β-helix domain, G244R destabilized the structure's free energy of folding by 10 kJ/mol, or 2.4 kCal/mol (Schuler and Seckler, 1998). These observations have led to the hypothesis that G244R and other *tsf* mutants act by globally destabilizing a monomeric folding intermediate where the β-helix domain is largely formed.

The global suppressor mutation (*su*) A334V is capable of rescuing an intrachain G244R folding defect or enhancing the folding yields of wild-type at higher temperatures *in vitro* and *in vivo* (Danner and Seckler, 1993; Fane et al., 1991; Mitraki et al., 1993). Although A334V does not alter the melting temperature of the full length trimer (Mitraki et al., 1993), it has been shown to accelerate the unfolding of ΔN and full length tailspike (Beissinger et al., 1995; Miller et al., 1998b). The crystal structure of A334V displays a root mean square difference of only 0.104 Å in comparison to the wild-type structure, but is hindered in its enzymatic activity by 40% (Baxa et al., 1999). This single amino acid mutation appears to stabilize the monomeric folding intermediate along the productive folding pathway, allowing for more efficient folding (Beissinger et al., 1995). Since A334V stabilizes the isolated β-helix domain by 7 kJ/mol, or 1.7 kCal/mol (Schuler and Seckler, 1998), it has been proposed that this suppressor mutation globally stabilizes the β-helical, monomeric folding intermediate.

3. Experimental Design

In previous chapters, it was shown that single alanine mutants of β-helix domain residues do not destabilize the native state of the trimeric P22 tailspike protein. Instead, mutant chain solubilities indicate the β-helix mutants affect a folding step before oligomerization. Do single alanine mutations located in the buried stacks of the tailspike β-helix domain prevent tailspike formation by hindering the folding of the β-helix?

To test the affects of mutations directly on the folding of the β-helix, mutations were constructed in the context of the isolated β-helix domain and assayed for their ability to fold *in vivo* at the temperatures previously tested with the full length tailspike. The mutations A334V and G244R were included in this study as controls.

B. MATERIALS AND METHODS

1. Production of Isolated β-helix Construct

pET(Δ*N*), which is a pET11a vector encoding a P22 tailspike protein lacking its N-terminal domain (1-108) called ΔN, was a generous gift from Dr. Anne Robinson (Gage and Robinson, 2003). Methylated pET(Δ*N*) DNA was purified to act as template DNA for mutagenesis from *E. coli* XL1-Blue cells (Stratagene) using a QIAprep Spin MiniPrep kit (QIAGEN).

The pET(Δ*N*) vector was mutated to a pET(*βHX*) vector, a pET11a vector encoding the isolated β-helix domain (residues 109-544) with a hexahistidyl tag at its C-terminus (Miller et al., 1998a), through a series of 3 site-directed mutagenesis reactions. The first reaction mutated three amino acids at the end of the β-helix domain of ΔN to histidines to generate ΔN-L545H-G546H-N547H, which was called ΔN-3H. The second reaction further mutated three more residues to hisitidines, generating ΔN-L545H-G546H-N547H-I548H-R549H-A550H, or ΔN-6H. The final reaction mutated the codons encoding residues 551 and 552 to amber and ochre stop codons, respectively, generating the βHX construct.

At each step of the process, methylated template DNA was mutated through a PCR-based QuikChange procedure (Stratagene) that utilized the appropriate mutagenic primer (IDT) listed in Table 4-1 and that primer's exact reverse complement. PCR products were transformed into *E. coli* XL1-Blue cells using a single step protocol (Chung et al., 1989). Following clonal growth, methylated plasmid DNA was purified as before and used either as a template for further rounds of mutagenesis or for DNA sequencing. DNA sequencing of a 400 base pair region (20% of the tailspike gene) containing the targeted codon confirmed the generation of ΔN-3H, ΔN-6H, and βHX clones (MGH DNA Core, Cambridge, MA).

2. Production of Single Alanine Mutants in the Isolated β-helix

Methylated pET(*βHX*) DNA was purified to act as template DNA for mutagenesis from XL1-Blue cells (Stratagene) as described above. Scanning alanine mutagenesis was performed using the PCR-based QuikChange mutagenesis procedure (Stratagene) as described in Chapter 2, Section B1. Mutagenic primers used to generate single alanine mutants in Chapter 2, Section B1 were used again to introduce the desired six mutations [F160A, F308A, F336A, F352A, Q489A, and C496A] individually into the βHX construct (Table 2-1). The mutagenic primers (IDT) listed in Table 4-2 and their reverse complements were used to generate the control mutations of G244R and A334V in the βHX-encoding gene. PCR products were transformed into *E. coli* XL1-Blue cells as described previously.

Plasmids encoding potentially altered βHX genes were purified from XL1-Blue cells and confirmed by sequencing, as described above. Plasmid DNA encoding confirmed alanine mutants was transformed into *E. coli* BL21(DE3) as described above. Cultures were selected on individual Luria Broth (LB) plates containing 100 μg/ml ampicillin. Cultures were made into 8.3% DMSO freeze stocks after 8 to 11 hours of growth in LB containing 100 μg/ml ampicillin at 37ºC. Freeze stocks were stored in dram vials at -80ºC.

Table 4-1. Mutagenic primers used to construct the isolated β-helix. Forward mutagenic primers are given that were used to produce the specified mutations. Reverse primers used in the reaction were the exact reverse complement of the forward primers.

Mutation	Forward primer
ΔN → ΔN-3H	5'-GGCAGAAGAAGGGCACCATCATATCCGCGCTAATAGTTTCGGC-3'
ΔN-3H → ΔN-6H	5'-GGGCACCATCATCATCACCACAATAGTTTCGGCTATGATAGC-3'
ΔN-6H → βHX	5'-TTGGCAGAAGAAGGGCACCATCACCATCACCACTAGTAATTCGGCTATGA-3'

Table 4-2. Mutagenic primers used to construct G244R and A334V. Forward mutagenic primers are given that were used to produce the specified mutations. Reverse primers used in the reaction were the exact reverse complement of the forward primers.

Mutation	Forward primer
G244R	5'-CGTTAAATTCCCACGAATAGAAACGTTACTCCCACC-3'
A334V	5'-GGGTCAGTAAGTAGCGTCCAGTTTTTACGTAATAATGGTGGC-3'

3. Production of G244R and A334V Full Length Tailspike Mutants

The plasmid pET(*gene9*) was mutated to encode the mutation G244R or A334V using the procedures described in Chapter 2, Section B1. The mutagenic primers listed in Table 4-2 were used to generate the G244R and A334V mutations. Mutants were confirmed by sequencing following transformation into XL1-Blue cells and miniprep of clonal cultures as described in Section B2. Confirmed mutant plasmids were transformed into *E. coli* BL21(DE3) cells and stored as 8.3% DMSO freeze stocks at -80ºC, as detailed in Chapter 2, Section B2.

4. Expression of β-helix and Full Length Constructs

Freeze stocks of full length tailspike constructs and pET(*βHX*) strains were streaked out on LB plates containing 100 μg/ml ampicillin and grown overnight at 37ºC. Clones from these plates or streaks from freeze stocks were used to inoculate 37ºC overnight growth cultures in LB containing 100 μg/ml ampicillin. Expression cultures were inoculated via 100-fold dilution of the overnight cultures into LB media containing 0.5% glucose, 50 mM MOPS and 100 μg/ml ampicillin. Cells were grown shaking (225 rpm) in 500 μl volumes in 96 deep-well plates for approximately 3.25 hours at 37ºC. IPTG was added to a final concentration of 1.5 mM to induce protein expression. Expression cultures were transferred to 37ºC, 30ºC, 18ºC, or 10ºC and were allowed to express protein for approximately 4, 8.5, 22.4 hours, or 24 hours respectively. Cultures were harvested via centrifugation for 15 minutes at 4000*g* and 4ºC. Cells were resuspended to 1/10 volume in lysis buffer (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA; 0.1% Triton X-100) and stored at -20ºC.

5. Cell Lysis, Fractionation, and SDS-PAGE

Frozen expression samples were thawed at room temperature and lysis was initiated by addition of freshly dissolved lysozyme (1 mg/ml final concentration from 10 mg/ml aqueous solution) and PMSF (9 mM final concentration from 90 mM stock in isopropanol). After 30 min of slow stirring at room temperature and addition of DNase I

(Sigma) (10 or 125 μg/ml final concentration from 0.06 or 1 mg/ml stock in glycerol; 1 M $MgSO_4$), lysates were stirred for an additional 30 min.

Half of each lysate was removed as a complete lysate sample, while the remaining lysate was centrifuged for 30 to 50 min at 4,000*g* and 4ºC in shallow well plates to separate soluble and insoluble fractions. Lysate pellets were resuspended in Buffer B (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA) by gentle agitation. Lysate fractions were mixed with 2% SDS sample buffer and run on 7.5% Triple Wide SDS polyacrylamide gels as described in Chapter 2, Section B3. Crude lysates, supernatant fractions, and insoluble pellet fractions were all analyzed by SDS-PAGE.

6. Optical Densitometry

Coomassie-stained gels were individually scanned at 400ppi and 8-bit gray scale via a ScanMaker 9800XL tabletop scanner (Microtek). Optical densitometry of the gel images was performed using the GelAnalyzer plugin (Appendix A) for ImageJ (NIH) as described in Chapter 2, Section 4. Background subtraction using lanes of expression cultures containing pET11a vectors lacking any tailspike or β-helix genes was performed separately for crude lysate samples, supernatant fraction samples, and pellet fraction samples.

Folding efficiencies for full length tailspike samples were calculated as described in Chapter 2, Section 4. For βHX constructs, the percent folded values were calculated as the percent ratio of the optical density of the supernatant fraction's βHX band versus the complete crude lysate's βHX band, according to the following formula:

$$\text{Percent Folded} = \frac{\text{OD}_{\text{Supernatant}}}{\text{OD}_{\text{Crude}}} * 100\%$$

These folding efficiencies were normalized to the average wild-type β-helix efficiency, as in Chapter 2, section 4.

C. RESULTS

1. Construction of ΔN-3H, ΔN-6H, and β-helix Constructs

Although the β-helix construct had been previously generated in a derivative of the pASK40 plasmid (Miller et al., 1998a), it was necessary to reproduce this construct in the context of the pET11a vector to achieve expression levels consistent with experiments discussed previously in this thesis. The plasmid pET(Δ*N*) was sequentially mutated to encode ΔN-3H, ΔN-6H, and, finally, the isolated β-helix (βHX) as detailed in section B1.

Once sequencing confirmed these three constructs, they were expressed at 37ºC and analyzed for their solubility and ability to fold. While expression of βHX predominantly produced soluble 48kDa polypeptide chains, ΔN-3H and ΔN-6H produced 72kDa chains that were found solely in the insoluble fraction of cell lysates. This data indicates that histidines in the region of 545 – 550 promote aggregation of the full length tailspike, but do not hinder the folding and solubility of the βHX protein when present as a C-terminal his-tag.

Six mutations investigated in the context of the full length tailspike were chosen for investigation in the context of the βHX construct: F160A, F308A, F336A, F352A, Q489A, and C496A. These mutations represent a diversity of folding phenotypes as determined in Chapter 2, of locations in the β-helix, and of amino acid types in the wild-type sequence (Figure 4-1). In addition to these six mutations, G244R and A334V were chosen as folding controls, since these mutations have been previously characterized in the full length tailspike protein and in the isolated β-helix, as discussed in Section A2. The eight desired mutants were generated in the context of the βHX construct, confirmed by sequencing, and subsequently assayed for their affect on the solubility of the βHX construct. Since G244R and A334V were not analyzed in Chapter 2, full length tailspike mutants were made, confirmed, and assayed for folding.

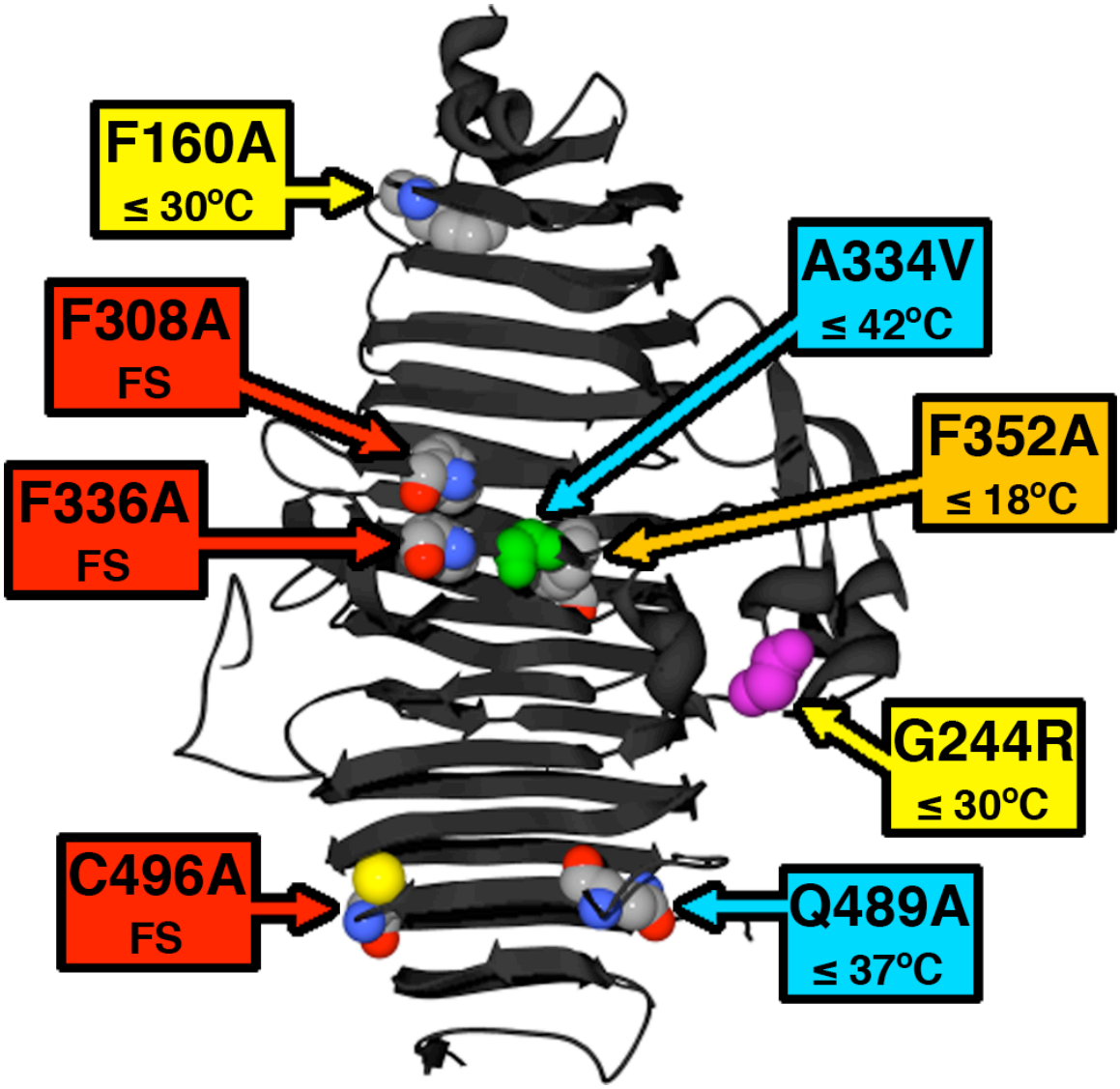

Figure 4-1. Mutations investigated in the isolated β-helix. The locations and folding phenotypes for the 8 mutations investigated in this chapter are shown on the structure of the β-helix domain from the native, trimeric structure (1TYU, (Steinbacher et al., 1996)). Folding phenotypes are given as the temperatures at which the mutations are wild-type-like folders *in vivo* in the context of the full length tailspike protein, along with a corresponding color code. Mutants that cannot fold at any temperatures tested are part of the folding spine and are labeled "FS". Folding temperatures for A334V and G244R are from the literature discussed in the main text.

2. *In Vivo* Folding Efficiencies of Full Length Tailspike Mutants

The newly constructed mutants A334V and G244R were over-expressed at all four temperatures previously tested in Chapter 2: 37ºC, 30ºC, 18ºC, and 10ºC. Since a number of the mutants under examination in this chapter are not in the folding spine and were not assayed at 10ºC, all 9 full length tailspike strains under analysis in this chapter were over-expressed at 10ºC. When unheated over-expression lysates were examined by SDS-PAGE, host protein translational levels were consistent across all samples and full length tailspike strains accumulated tailspike polypeptide chains as clearly resolved bands. These observations are consistent with those made in previous expression experiments in Chapter 2.

The percent of wild-type folding efficiencies were calculated from three or six independent expression samples for each strain at each temperature newly tested. Without normalization, the folding efficiency of wild-type chains expressed at 10ºC was 95±3% (3 independent expression samples), which is consistent with the previously determined value of 93±4%. This consistency was also observed for mutant strains previously examined at 10ºC, F308A, F336A, and C496A (Table 2-2, Table 4-3). At most, new and old values differed by approximately 9%.

As had been observed previously, all strains having a folding efficiency less than 75% of wild-type at 18ºC were folding deficient at 10ºC, while mutants that folded like wild-type at 18ºC continued to do so at 10ºC (Table 4-3). As expected from previous publications, A334V folded like wild-type or better at all temperatures tested while G244R only displayed greater than 75% folding efficiencies at 30ºC or lower temperatures. Only 2±2% of G244R tailspike chains reached the native state at 37ºC, which is consistent with the original discovery that G244R is a temperature sensitive folding mutant.

Table 4-3. Folding efficiencies of full length tailspike mutants. The percent of WT folding efficiencies for 8 mutants and WT at the four temperatures tested are given as averages and standard deviations of multiple expression experiments. Folding deficient average values (<75%) are shaded and bolded, while new expression data is starred (*).

Sample	Percent of WT folding efficiency			
	37ºC	30ºC	18ºC	10ºC
WT	100.0 ± 9.9	100.0 ± 5.7	100.0 ± 1.8	100.0±3.2 *
F160A	**8.8 ± 4.9**	77.4 ± 2.7	100.4 ± 0.8	103.6±1.3 *
G244R	**2.2±1.6 ***	81.9±5.9 *	102.6±11.6 *	98.7±5.0 *
F308A	**0.0 ± 0.0**	**0.7 ± 1.0**	**8.4 ± 10.7**	**0.7±0.7 ***
A334V	109.9±26.7 *	98.6±3.2 *	105.7±2.4 *	100.9±1.3 *
F336A	**0.0 ± 0.0**	**0.1 ± 0.1**	**28.5 ± 32.8**	**47.1±1.9 ***
F352A	**0.6 ± 0.9**	**11.7 ± 16.1**	95.8 ± 4.5	96.4±1.9 *
Q489A	102.2 ± 11.9	100.7 ± 4.2	92.6 ± 7.2	90.5±3.0 *
C496A	**0.7 ± 1.0**	**20.4 ± 19.4**	**58.5 ± 13.4**	**6.3±2.9 ***

Table 4-4. Folding efficiencies of βHX mutants. The percent of WT-βHX folding efficiencies for 8 mutants and WT-βHX at the four temperatures tested are given as averages and standard deviations of triplicate expression experiments. Folding deficient average values (<75%) are shaded and bolded.

Sample	Percent of WT-βHX folding efficiency			
	37ºC	30ºC	18ºC	10ºC
WT-βHX	100.0±22.1	100.0±18.8	100.0±10.2	100.0±5.9
F160A-βHX	**0.8±0.2**	121.8±6.2	102.7±5.0	97.1±3.5
G244R-βHX	**0.5±0.4**	79.9±4.6	76.7±5.4	84.9±5.7
F308A-βHX	**0.3±0.1**	**0.0±0.1**	**0.0±0.0**	**0.0±0.1**
A334V-βHX	116.4±3.9	115.0±11.3	84.1±2.5	105.4±6.0
F336A-βHX	**0.0±0.0**	**0.1±0.1**	**8.6±3.1**	**49.0±8.5**
F352A-βHX	**0.1±0.1**	**0.0±0.0**	**67.4±6.4**	85.2±9.9
Q489A-βHX	119.3±5.4	119.1±12.2	107.2±14.5	89.5±1.6
C496A-βHX	**0.0±0.0**	**2.7±2.2**	96.3±1.0	**12.0±3.8**

3. *In Vivo* Expression of β-helix Mutants

Mutant and wild-type βHX proteins were expressed in *E. coli* cultures at 37ºC, 30ºC, 18ºC, and 10ºC in a manner identical to the full length tailspike proteins. As observed with the full length tailspike samples, host protein translational levels were consistent across all samples when examined by SDS-PAGE (Figure 4-2). The SDS-sensitive βHX protein band had a mobililty corresponding to its 48kDa molecular weight and could be identified in all samples by comparison with a pET11a expression sample, which lacked any tailspike or β-helix genes (Figure 4-2A). Examination of the crude lysates of all expression samples indicated that the βHX protein was expressed in a consistent manner and was not degraded by host proteases at any of the temperatures tested.

When crude lysates were fractionated by low speed centrifugation, βHX proteins partitioned between the soluble supernatant fraction and the insoluble pellet fraction. Based on previously published fluorescence and CD spectroscopy, oligosaccharide binding, and enzymatic activity of purified, soluble β-helix constructs expressed at 30ºC, which include wild-type, G244R, and A334V, it is assumed that solubility is a measure of folding *in vivo* (Miller et al., 1998a; Schuler and Seckler, 1998). The observed partitioning and consistent expression levels allowed for the quantitation of folding efficiencies for all isolated β-helices. The most reliable method for calculating the percent of βHX chains folded proved to be a ratio of the supernatant chains to crude lysate chains, due to concentration changes that occur in resolubilizing the pelleted fraction. Since this calculation requires the comparison of two distinct gel electrophoresis samples while the calculation for full length tailspike relies only on a single crude lysate sample, this method inherently introduces more potential sources of error. In spite of this potential error, the calculated folding efficiencies proved to be both precise and consistent with previous experiments (Table 4-4, Table 4-3).

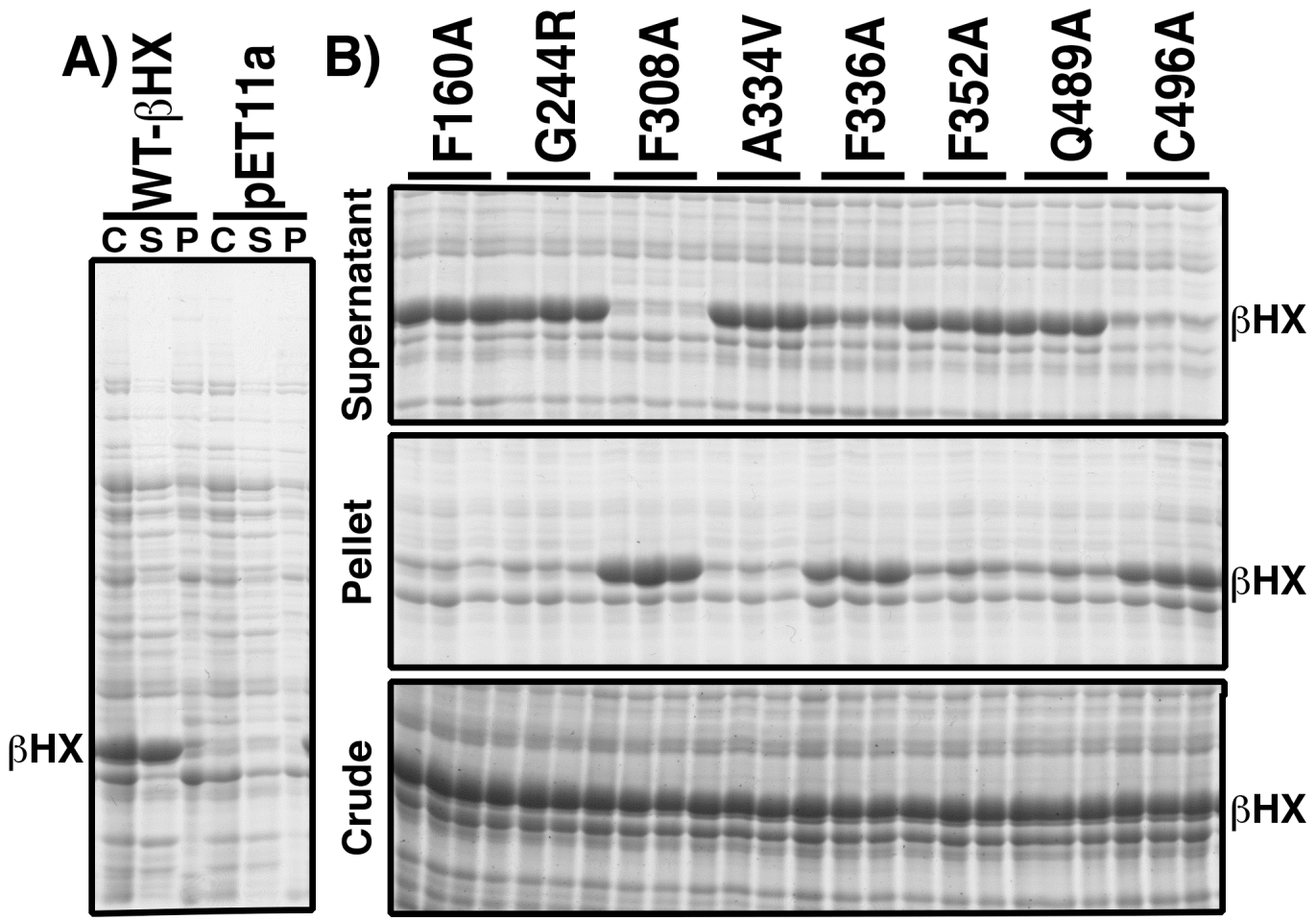

Figure 4-2. Sample SDS-PAGE of β-helix samples. Crude lysates (C), supernatant fractions (S), and pellet fractions (P) from 10ºC overnight expression cultures were analyzed by SDS-PAGE for β-helix expression and solubility. The locations of the 48kD β-helix band are marked as βHX. (A) SDS-PAGE of wild-type βHX and pET11a, which lacks any βHX or tailspike genes. (B) Triplicate expressions of eight single mutations in the isolated β-helix construct.

4. Wild-type, G244R, and A334V Folding Efficiencies

The calculated folding yields for the wild-type isolated β-helix proteins demonstrate some of the errors in calculating folding yields from multiple samples. For example, when wild-type βHX was expressed at 18ºC in three independent expression cultures, the folding yield was 127±13%. Folding yields significantly greater than 100% are indicative of some error in fractionation or pipetting. The calculated folding yields of expression at 37ºC, 30ºC, and 10ºC, 105±23%, 88±17%, and 88±5% respectively, did not appear to suffer from these errors, and it can be observed that the wild-type β-helix consistently produced high folding yields at all temperatures tested.

Normalization of mutant folding yields to the wild-type βHX folding yields at each temperature generated folding efficiencies for A334V and G244R that are consistent with the folding efficiencies of full length tailspike expression cultures (Figure 4-3) and previously published data. A334V-βHX was able to fold like wild-type or better at all temperatures tested (Table 4-4), consistent with the folding efficiencies observed for the full length A334V tailspike (Table 4-3). Although the folding efficiencies of A334V-βHX do not always overlap with those of full length A334V, they are never more than 22% apart (Figure 4-3).

G244R-βHX accumulated in the pelleted fraction of the lysate when expressed at 37ºC, yielding a relative folding efficiency of 0.5±0.4%. Below 37ºC, G244R-βHX folded with wild-type-like folding efficiencies (Table 4-4). This temperature sensitivity is consistent with folding efficiencies obtained with the full length G244R tailspike (Table 4-3, Figure 4-3). The wild-type-like folding of G244R-βHX at 30ºC reflects previously published data, where this construct was expressed at 30ºC in generating soluble protein for purification (Schuler and Seckler, 1998).

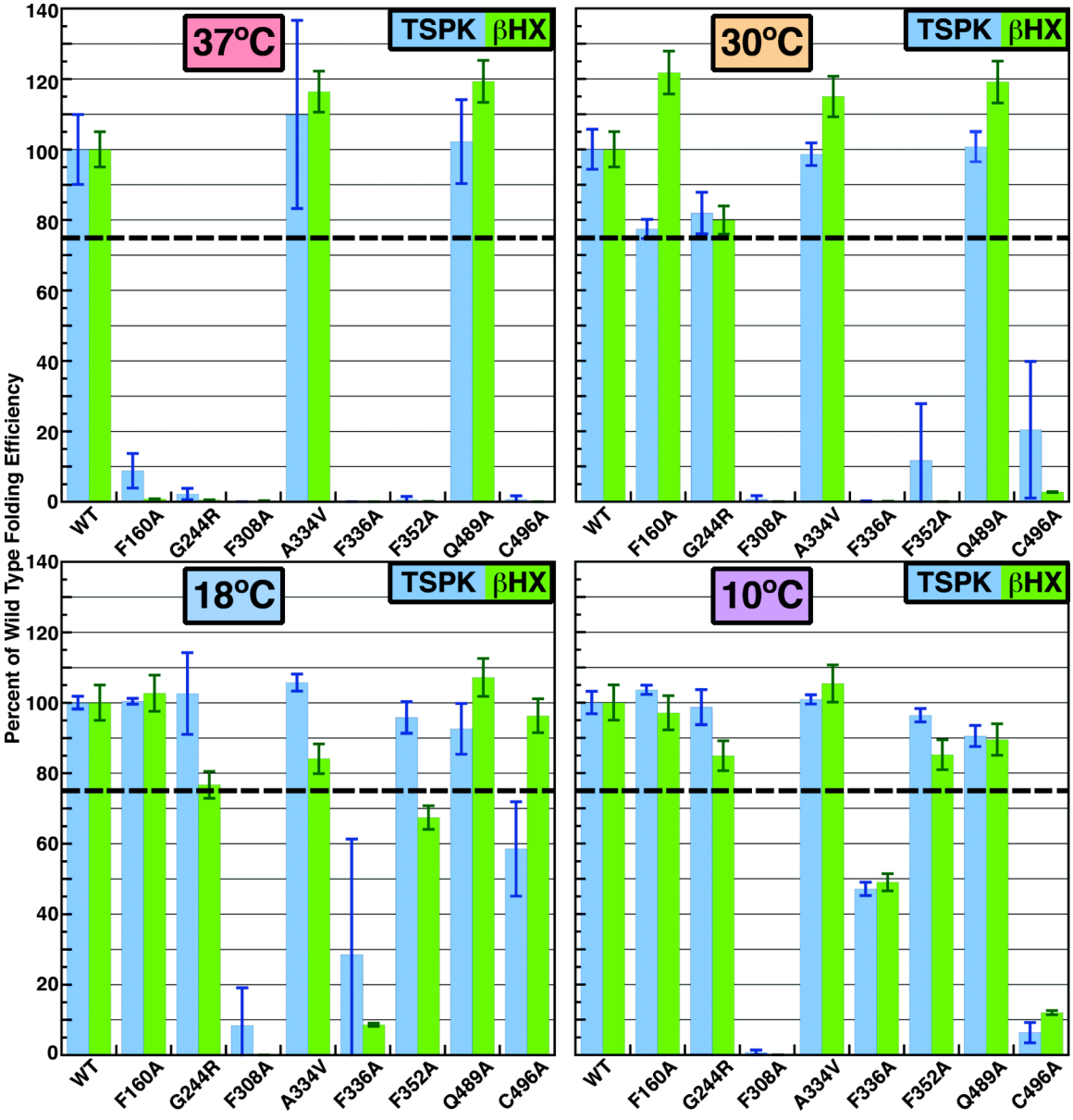

Figure 4-3. Quantitative comparison of β-helix and tailspike *in vivo* folding. Average percent of wild-type folding efficiencies are shown for full length tailspike (blue) and βHX (green) mutants at the four temperatures tested. Darker color lines indicate the standard deviations of triplicate experiments. The 75% threshold delineating wild-type-like folders and folding deficient samples is shown as a dashed line.

5. Single Alanine Mutant βHX Folding Efficiencies

The folding efficiencies for single alanine mutations of the isolated β-helix reflect the observations made with A334V and G244R that the solubility of βHX constructs matches the folding efficiencies of full length tailspike mutants. At 37ºC, the only alanine mutant that is capable of folding with wild-type-like folding efficiencies is Q489A-βHX (Table 4-4). All other alanine mutants are completely incapable of yielding soluble βHX proteins. This is identical to the observed folding efficiencies of full length constructs at 37ºC (Figure 4-3).

When expressed at 30ºC, F160A was the only alanine mutant incapable of folding at 37ºC that was rescued to a wild-type-like folding efficiency. As observed with the full length tailspike proteins, all other alanine mutants failed to fold and accumulated in the pelleted fractions of the cellular lysates.

At 18ºC, F160A-βHX, Q489A-βHX, and C496A-BHX folded with wild-type-like efficiencies, while F308A-βHX, F352A-βHX, and F336A-βHX were folding deficient. The folding efficiencies of F160A, F308A, and F336A are nearly identical to those calculated for the full length tailspike protein (Table 4-3, Table 4-4, Figure 4-3). For Q489A, both full length and βHX proteins are wild-type-like in folding but show a 14.6% difference. F352A-βHX is folding deficient with a folding efficiency of 67±6% while F352A full length is classified as a wild-type-like folder with a folding efficiency that is 28% higher than the βHX construct. Although there is a discrepancy in the folding efficiencies of the full length and βHX F352A, it is clear that F352A-βHX is capable of generating a substantial amount of soluble proteins. The same is true for C496A, where the full length tailspike is folding deficient at 18ºC, but C496A-βHX displays a wild-type-like folding efficiency 37.8% greater than its full length counterpart.

Isolated β-helix mutants F308A, F336A, and C496A were folding deficient when folded *in vivo* at 10ºC. The calculated folding efficiencies for these mutants are nearly identical to those observed for their corresponding full length tailspike mutants. This is also the case for F160A, F352A, and Q489A, which are wild-type-like folders in the

context of the isolated β-helix and in the full length tailspike protein. It is interesting to note that C496A, both in the isolated β-helix domain and in the full length protein, is capable of producing a substantial amount of folded protein at 18ºC but is nearly completely blocked in folding at 10ºC.

D. DISCUSSION

1. Alanine Mutants Affect the Folding of the β-helix

Seckler and coworkers were able to purify and characterize wild-type and mutant isolated β-helix domain proteins from the soluble fractions of 30ºC expression cultures. Based on a number of biophysical methods, they concluded that these proteins were in a β-helical confirmation that is very close to the native β-helix conformation observed in the full length protein (Miller et al., 1998a; Schuler and Seckler, 1998). In the experiments presented here, the *in vivo* solubilities of mutant and wild-type βHX proteins mirrored the corresponding full length tailspike sequences' abilities to produce native, SDS-resistant trimers in the cellular environment. Taken together, these observations support the notion that solubility is a measure of the folding of the βHX domain *in vivo*.

Overall, the folding phenotypes observed here for the isolated β-helix domain reasonably replicate the folding efficiencies observed for the full length tailspike mutants. Two mutants did not follow this general observation at one temperature, 18ºC. Although these βHX mutants crossed the 75% delineation between folding deficiency and wild-type-like folding with respect to their corresponding full length constructs, both full length and βHX constructs produced substantial amounts of folded protein. In contrast, mutant F308A was never able to produce any folded protein at any temperature in either the full length protein or the isolated β-helix domain. The data for all mutants across all four temperatures suggests that the single alanine mutants affect the ability of the isolated β-helix to fold. This in turn demonstrates that the alanine mutants in the buried core of the β-helix domain of the P22 tailspike protein prevent the full length tailspike from folding by preventing the β-helix domain from folding.

2. Multiple Pathways of Aggregation

In constructing the pET(*βHX*) construct, it was noted that ΔN-3H and ΔN-6H were completely blocked in forming any native, trimeric tailspike proteins and instead accumulated in the insoluble fraction of the cellular lysate. In contrast, βHX predominantly accumulated in the soluble fraction, indicating that it was folded as investigated by Seckler and co-workers (Miller et al., 1998a). Since ΔN-6H and βHX differ from each other by the presence of the C-terminus of the tailspike protein (residues 551 – 666), this region of the protein must be responsible for the aggregation observed with the ΔN-6H construct.

Various mutants of βHX accumulated in the pelleted fraction of crude lysates in a temperature dependent manner. One particular mutant, F308A, was never capable of producing any soluble βHX proteins at any temperatures and consistently produced insoluble aggregated protein. Previous investigations into the aggregation of the isolated β-helix domain have shown that this protein fragment is not a well-behaved protein and readily aggregates (Miller et al., 1998a; Schuler et al., 1999).

The aggregation of the ΔN-6H and the aggregation of the isolated β-helix domain depend upon different parts of the full length tailspike protein. Previous investigations into the aggregation pathway and inclusion body formation of tailspike proteins indicate that aggregation proceeds from a partially folded or misfolded β-helical intermediate that is aggregation prone (Haase-Pettingell and King, 1988; King et al., 1996; Schuler et al., 1999). The results described here demonstrate that multiple pathways of aggregation exist for the P22 tailspike protein (Figure 4-4). One pathway utilizes the C-terminal, oligomerization domain while the second involves the β-helix domain. It is unclear why these distinct pathways have not been detected before, but it may be postulated that the process of evolution optimized the C-terminal region for folding, assembly, and stability, while it has been previously demonstrated that the β-helix domain has been optimized for oligosaccharide binding rather than folding and stability (Seckler, 1998). Hence, in the full length protein, the region less optimized for folding is more likely to cause aggregation.

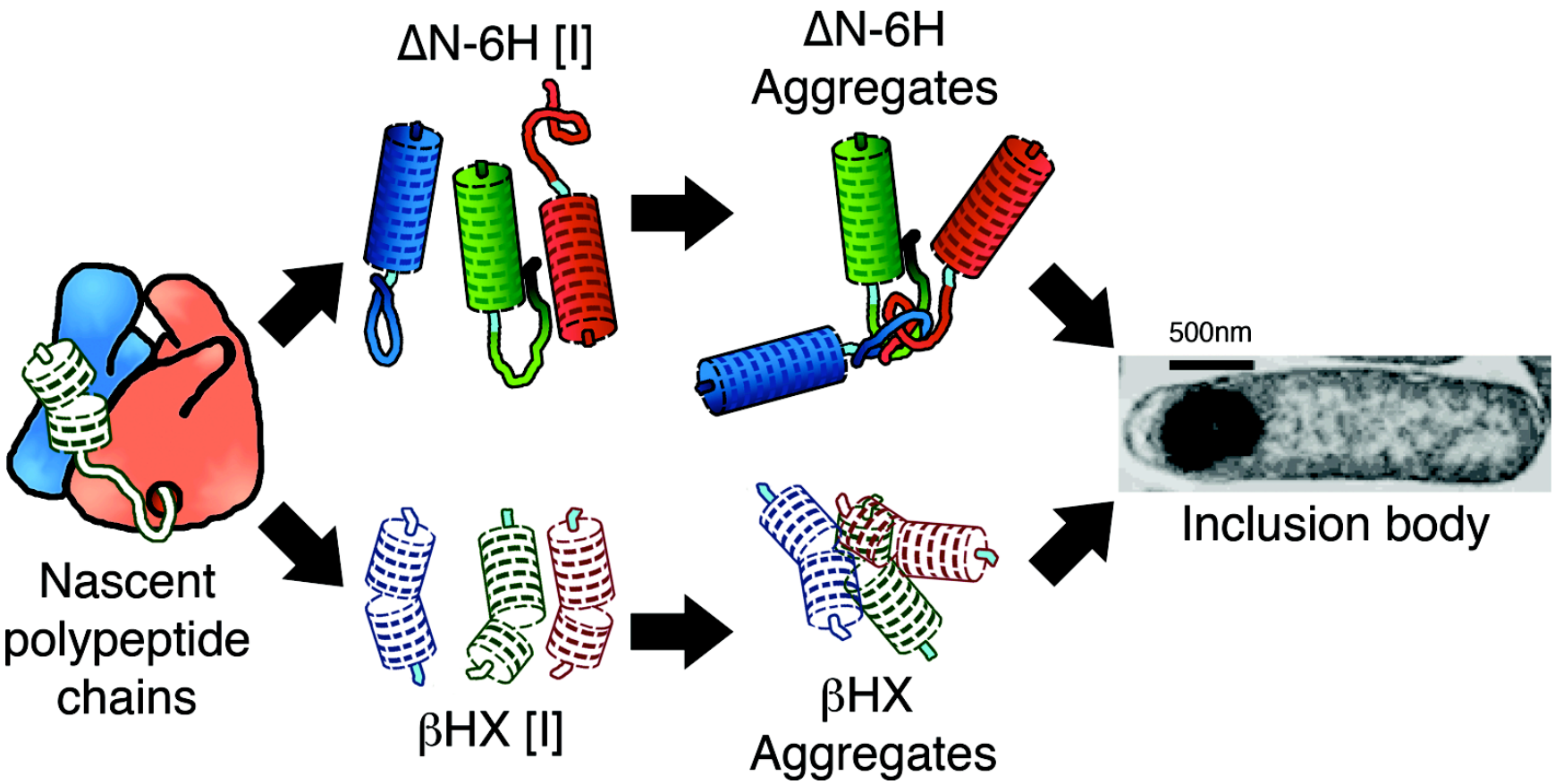

Figure 4-4. βHX and ΔN-6H aggregate through different pathways. When released from the ribosome, ΔN-6H likely folds to an intermediate structure where the β-helix domain is folded, but the C-terminal part of the protein is not. Since the hexahistidyl internal tag (cyan) of ΔN-6H appears to block folding, this protein proceeds to inclusion body accumulation through aggregation of the C-terminal tails of ΔN-6H proteins. In contrast, the βHX construct likely folds into a partially folded intermediate at high temperatures or in the presence of folding defective mutants. This aggregation prone intermediate oligomerizes and accumulates into the inclusion body.

CHAPTER FIVE:

IN VITRO REFOLDING OF TAILSPIKE MUTANTS

A. INTRODUCTION

1. *In vitro* refolding of P22 Tailspike

In vitro refolding of the tailspike protein has shown that the protein is capable of refolding without the presence of the ribosome or other cellular factors and that *in vitro* refolding closely resembles the intracellular maturation pathway (Fuchs et al., 1991). Therefore, cellular components are not necessary for the folding of the protein and the amino acid sequence is sufficient to encode the folding and assembly of the protein to its native, trimeric state.

Tailspike folding *in vitro* is rarely as efficient as *in vivo*. For example, folding yields as high as 60% have been reported for *in vitro* refolding of the wild-type tailspike at 10ºC (Fuchs et al., 1991; Seckler et al., 1989), while a 10ºC *in vivo* folding yield of 93% was achieved in Chapter 2 of this thesis. One difference between the *in vitro* and *in vivo* refolding pathways that may account for this lower yield *in vitro* is that nascent polypeptide chains on the ribosome exhibit a β-helical conformation, as determined by antibody binding studies, that is not present at early times when tailspike is refolded *in vitro* (Clark and King, 2001). This observation has led to the suggestion that interaction with the ribosome or other cellular factors assists the tailspike chain in reaching these productive folding conformations more quickly than upon refolding *in vitro*.

2. Chaperones and the P22 Tailspike

In the cell, tailspike protein folding intermediates are recognized and bound by GroEL, but folding defects are not rescued by the presence of the chaperone (Gordon et al., 1994). *In vitro* studies showed that although GroEL can bind refolding tailspike, the tailspike protein is not released in an efficient manner and can thereby be trapped when ATP is not present (Brunschier et al., 1993). In contrast, the coat protein of the P22 phage has been shown to interact with GroEL to suppress aggregation (Doyle et al., 2004). Secondary-site suppressors of temperature sensitive folding mutants in the coat protein suppress aggregation by enhancing interactions between folding intermediates of the coat protein and the chaperone system (Parent et al., 2004).

Although the tailspike does not utilize the GroEL chaperone, the nascent chain displaying β-helical conformation does associate with the 30S ribosomal subunit (Clark and King, 2001). Since the ribosome-associated chaperone trigger factor and other chaperones such as the signal recognition particle dock to the 50S subunit near the polypeptide exit tunnel (Kramer et al., 2002; Schaffitzel et al., 2006), it is unlikely that these factors direct the folding of this 30S bound conformer. It has been suggested that co-translational interactions with the ribosome may favor β-helix formation.

3. Experimental Design

The results presented in previous chapters demonstrate that alanine mutations in the buried core hinder the folding of the β-helix domain in a temperature-sensitive manner. Do these residues affect intermolecular interactions, such as with the ribosome or chaperones, that favor the folding of the β-helix, or are the wild-type residues sufficient for encoding the folding of the β-helix domain?

To investigate if the amino acid substitutions investigated in Chapter 4 are sufficient for producing the observed *in vivo* folding defects documented in Chapter 2, full length tailspike mutant polypeptide chains that had been investigated in the context of the isolated β-helix were purified, chemically denatured, and assessed for their ability to refold *in vitro*. Since mutant polypeptide chains completely incapable of reaching the native state at all temperatures were amongst those to be isolated, methods for purifying native tailspike trimers could not be utilized. Although his-tagged, full length tailspike proteins were constructed and purified, the additional amino acids present in the tag interfered with wild-type tailspike refolding (see Appendix B). Therefore, inclusion bodies from over-expressing cultures of full length tailspike sequences were purified and chemically denatured. Polypeptide chains were refolded *in vitro* to determine how alanine mutants in the β-helix domain affect the refolding of tailspike trimers.

B. MATERIALS AND METHODS

1. Production of Inclusion Bodies of Full Length Tailspike Mutants

Freeze stocks of full length tailspike constructs, which were generated as detailed in Chapter 2 and above, were streaked out on LB plates containing 100 μg/ml ampicillin and grown overnight at 37ºC. Clones from these plates were used to inoculate 37ºC overnight growth cultures in Super Broth containing 100 μg/ml ampicillin. Expression cultures were inoculated via 200-fold dilution of the overnight cultures into 1 liter Super Broth media containing 100 μg/ml ampicillin. Cells were grown shaking (225 rpm) for approximately 3 hours at 37ºC. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. Expression cultures were transferred to 42ºC water baths and were allowed to express protein for 5 hours. Cultures were harvested via centrifugation for 15 minutes at 4000*g*. Cell pellets were resuspended in 10 ml of lysis buffer (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA; 0.1% Triton X-100) and stored at -80ºC.

Frozen expression cultures were thawed at room temperature and lysed at 4ºC and 18,000 psi via French press. Samples were pressed three times each before adding PMSF (0.9 mM final concentration from 90 mM stock in isopropanol). Inclusion bodies were harvested by centrifugation at 5,800*g* for 40 minutes. Pelleted inclusion bodies were washed twice with 25 ml of lysis buffer for over an hour at 4ºC after being resuspended by scraping with a glass rod. Washed samples were pelleted by centrifugation at 15,000*g* for 10 minutes at 4ºC. Purified inclusion bodies were resuspended in 25 ml of Buffer B (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA) and stored at 4ºC for up to four months.

Inclusion body samples were analyzed by SDS-PAGE to observe their purity and to estimate the concentration of tailspike protein in each sample via optical densitometry.

2. Unfolding and Refolding of Inclusion Bodies

Inclusion body samples containing approximately 40 μg of tailspike protein were pelleted via centrifugation for 10 minutes at 10,000*g* at room temperature. Pellets were resuspended in unfolding buffer (6 M urea, 10 mM DTT, 50 mM Tris-HCl, 25 mM NaCl, 2 mM EDTA, pH 7.6) to an estimated concentration of 2mg/ml and incubated for 1 hour at room temperature.

Unfolded samples were rapidly diluted 1:20 with ice cold refolding buffer (50 mM Tris, 2 mM EDTA, pH7.6) to initiate refolding with an approximate tailspike concentration of 100 μg/ml. Samples were placed on ice for 30 minutes before being moved to a water bath at the desired testing temperature, either 37ºC, 30ºC, 18ºC, or 10ºC. After specified time points, aliquots were removed from the refolding sample and mixed with 2% SDS sample buffer and stored at 4ºC until all time point samples were taken and could be analyzed by gel electrophoresis.

3. SDS-PAGE and Optical Densitometry of Refolding Samples

Time point samples were electrophoresed through 7.5% polyacrylamide gels as described in Chapter 2, Section B3. Gels were fixed with 10% acetic acid and 30% ethanol for over one hour and then stained with SYPRO® Ruby stain (Invitrogen) overnight. Gels were visualized using the Typhoon 9400 variable mode imager (Amersham Biosciences) with a fluorescence excitation of 457 nm, an emission filter of 610 nm, and a photomultiplier value of 750 – 800. Gels were scanned at a resolution of 100 microns.

Optical densitometry of the gel image was performed using the GelAnalyzer plugin (Appendix A) for ImageJ (NIH). Images were calibrated using ImageJ's Uncalibrated OD correlation function. Profile plots calculated for sample lanes were not corrected by a background lane, but were corrected by an average baseline value. As described in Chapter 2, Section B4, integration of the positive area of peaks corresponding to tailspike bands yielded relative protein amounts. A percent folded was calculated for each time point as the percent of tailspike chains present in the native, trimeric tailspike band, as detailed in Chapter 2, Section B4.

4. Kinetic Fits of Time Course Data

Average percent folded time course data were fit to a single exponential rise function using the curve fitting feature of Kaleidagraph. This particular fit has the following formula, where t is time and A, B, and C are fitting variables:

$$\text{Percent Folded} = A + B(1 - e^{(-C*t)})$$

Each average percent folded value, calculated from triplicate independent experiments, was weighted by one over the square of the calculated standard deviations of triplicate experiments. Kaleidagraph provides data weighting as an option for curve fitting.

C. RESULTS

1. Inclusion Body Production

Full-length tailspike mutants that had been examined at 10ºC in Chapter 4 (Figure 4-1) were over-expressed at 42ºC, a temperature at which even the suppressor mutation A334V accumulates into inclusion bodies rather than folding or undergoing intracellular degradation. Inclusion bodies were washed and purified via several rounds of low speed centrifugation. Purified inclusion body samples were stable at 4ºC for several months with no signs of proteolysis.

All tailspike sequences produced inclusion bodies whose composition was 70 – 75% SDS-sensitive tailspike protein (Figure 5-1). In contrast, a pET11a lacking any tailspike genes was induced for expression but yielded no inclusion body proteins. Although some samples, such as A334V, still produced some native trimers at 42ºC, these trimers were removed in the process of purifying the inclusion bodies.

2. Inclusion Body Refolding

Pelleted inclusion body samples were chemically denatured at neutral pH in 6 M urea. Although these conditions are not strong enough to denature the natively folded, trimeric tailspike (Seckler et al., 1989), they were sufficient for solubilizing the inclusion

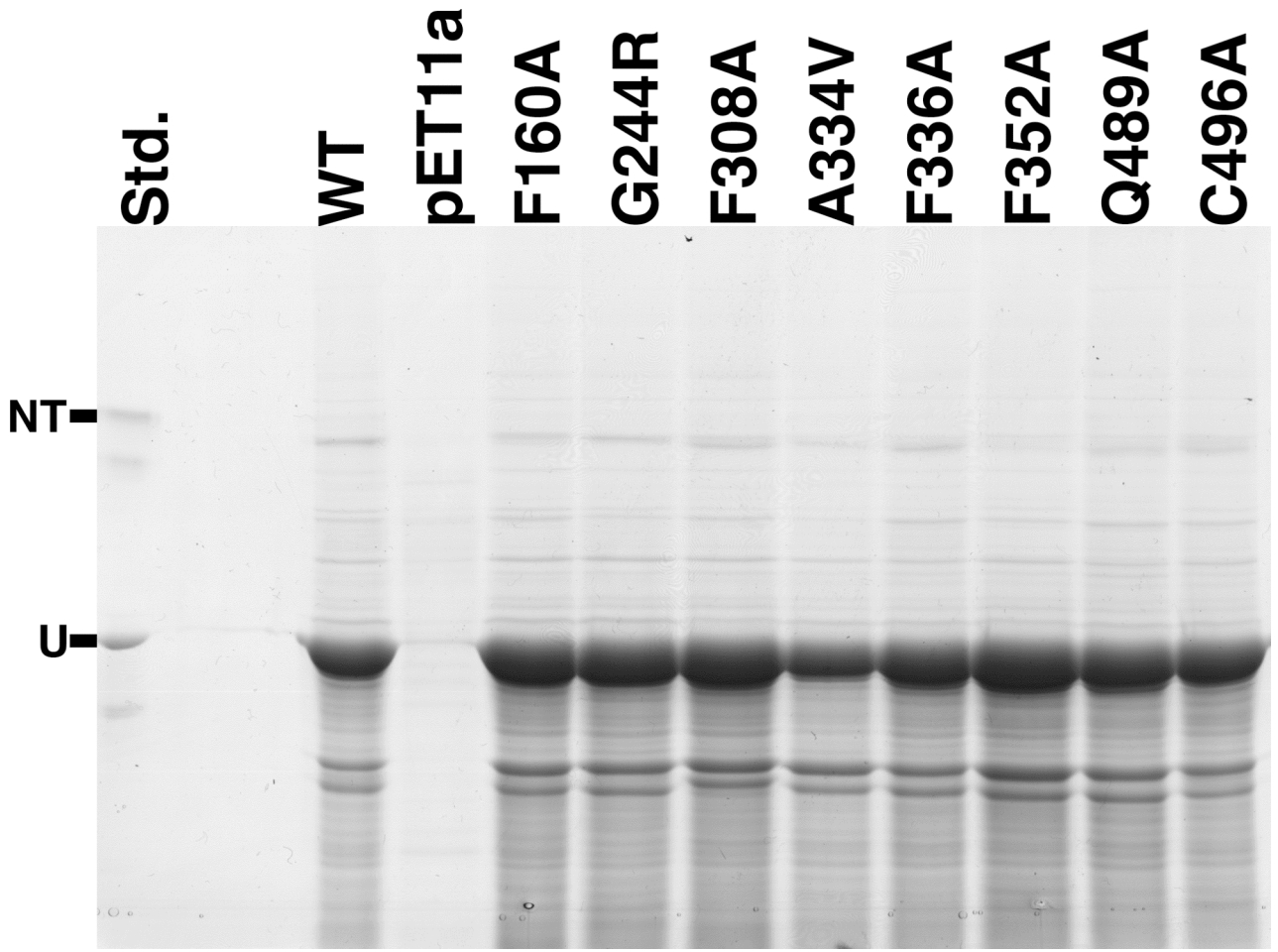

Figure 5-1. Purified inclusion body samples. *E. coli* cultures expressing wild-type (WT) tailspike, mutant tailspikes, and a pET11a plasmid lacking any tailspike genes were induced for expression at 42ºC. Inclusion body samples were purified and analyzed by SDS-PAGE. Tailspike specific bands can be identified by comparison with a purified tailspike standard (Std.) containing native trimers (NT) and unfolded SDS-sensitive polypeptide chains (U).

body pellets. After one hour of unfolding at room temperature, samples were diluted with ice cold refolding buffer and incubated for 30 minutes on ice. This ice step has been shown to suppress *in vitro* aggregation of folding intermediates and produce folding yields in excess of 80% at 20ºC (Betts and King, 1998). Since tailspike *in vitro* refolding yields are typically much lower than *in vivo*, and since a number of the mutations examined in this study have already proven to be highly aggregative and defective in folding, we wanted to refold under optimal conditions for productive refolding. After the incubation on ice, refolding time courses were monitored for wild-type and all mutant samples in triplicate at three different temperatures: 10ºC, 18ºC, and 30ºC. SDS gels of 18ºC refolding samples for wild-type, G244R, and F308A are shown in Figure 5-2.

As can be seen in Figure 5-2, all tailspike chains were in an SDS-sensitive state upon being moved from ice to a heated water bath. With increasing time of incubation, wild-type and G244R tailspike chains trimerized and became SDS-resistant. This was the case for a number of other mutants, as will be discussed below. As expected, the G244R trimer exhibited an altered electrophoretic mobility with respect to wild-type trimers. While WT and G244R accumulated as trimers, the density of the SDS-sensitive band (U) diminished with time, as is most obvious in the case of WT. In contrast to the ability of wild-type and G244R polypeptide chains to reach the native state, F308A chains were completely incompetent for folding and remained in the SDS-sensitive band throughout the refolding time course.

Except for two bands, marked by a question mark in Figure 5-2, bands representing other proteins in the inclusion bodies did not increase or decrease in density through the course of the experiment. The marked bands, however, increased sharply after the ice incubation and decreased with time. These bands appear to be tailspike specific, as they are not present in the refolding time course of the folding defective mutant F308A. The unknown band in the G244R sample is electrophoretically altered with respect to the unknown band in the wild-type sample, as would be expected if these bands were tailspike specific. Although these bands may be tailspike fragments, such as ΔN tailspike chains that are known to be produced by proteolysis of partially unfolded intermediates (Chen and King, 1991), the fact that their density decreases with time

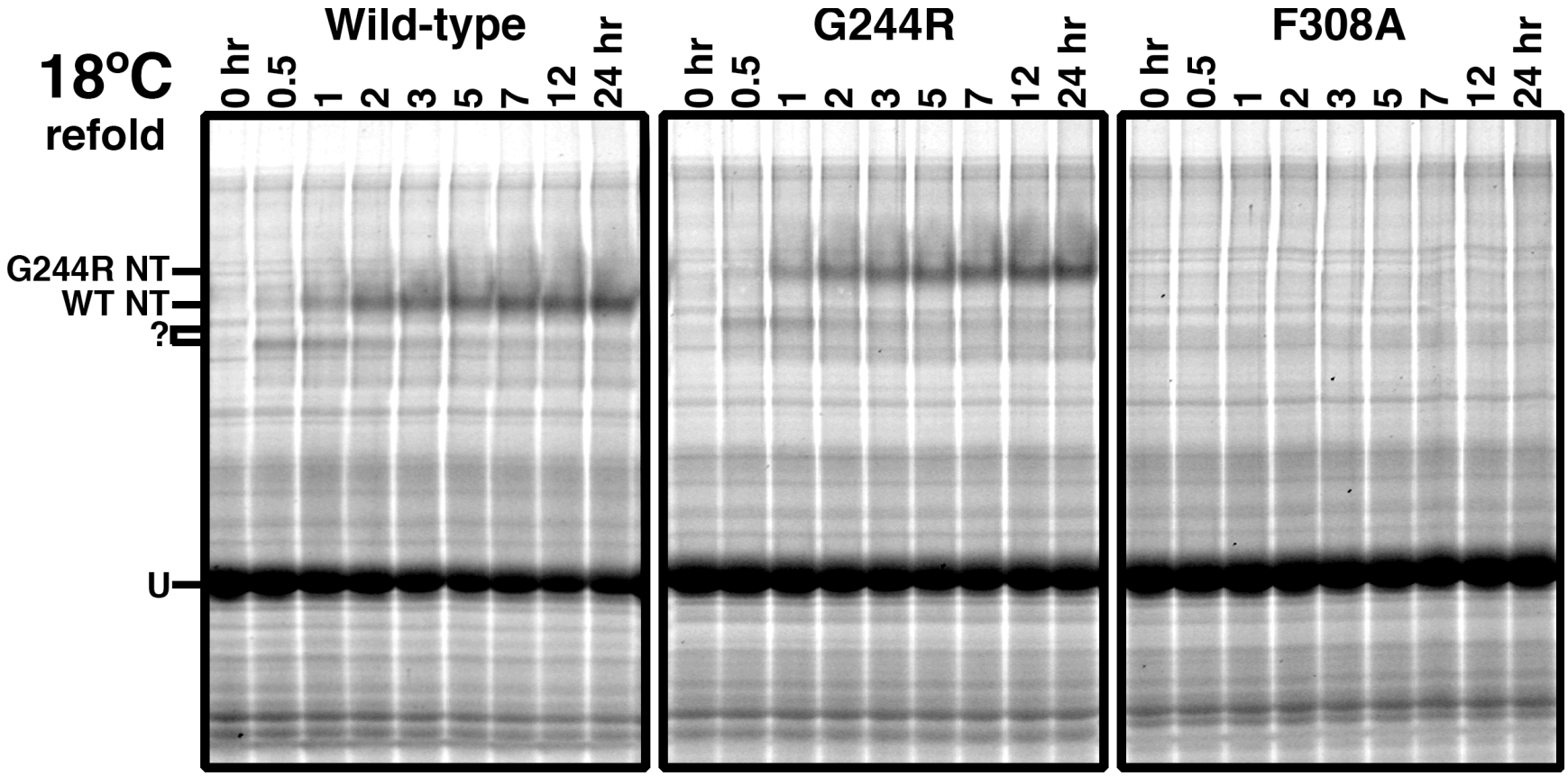

Figure 5-2. Sample SDS-PAGE of inclusion body refolding. 18ºC inclusion body refolding experiments were analyzed by SDS-PAGE. Samples incubated for varying times at 18ºC (given in hours) following an initial 30 minute ice incubation. All tailspike chains began as non-native, SDS-sensitive chains that unfold (U) upon mixture with SDS. With increasing time of incubation, wild-type chains folded to the SDS-resistant, native trimer state (WT NT). G244R SDS-resistant, native trimers (G244R NT) display an altered mobility, as discussed in previous publications (Goldenberg et al., 1982; Yu and King, 1984). A previously unidentified tailspike-specific band (?) appears in the wild-type and G244R samples beginning at 0.5 hours and diminishes with time.

suggests that they are previously unidentified, SDS-resistant trimeric folding intermediates. However, since it is not known if these bands represent full length tailspike proteins, the calculation of percent folded values at each time point did not include these bands.

3. Refolding Time Courses

Refolding data acquired from the SDS-PAGE analysis of refolding time points demonstrated that the mutations investigated here significantly affect the refolding of tailspike chains to the SDS-resistant trimer state. This is most readily apparent from the refolding time courses of C496A, F336A and F308A. These full length tailspike mutants were completely incapable of producing substantial SDS-resistant tailspike proteins above the values calculated at the zero hour time point, independent of the refolding temperature **(**Figure 5-3).

In contrast, the wild-type protein was capable of reaching refolding yields of 26±2% at 30ºC, 43±4% at 18ºC, and 29±1% at 10ºC. The time courses observed in Figure 5-3 could be fit by a single exponential rise function. For wild-type, the resulting fits had coefficients of determination (R^2) of 0.98 at 30ºC, 0.98 at 18ºC, and 0.99 at 10ºC, indicating the goodness of the fits. These fits were used to determine the half-times of refolding, which are shown for all time courses in Table 5-1. For the wild-type tailspike, the calculated half times of refolding, 12 minutes at 30ºC, 85 minutes at 18ºC, and over 4 hours at 10ºC are generally in agreement with previously published data, which cites the refolding half-times of wild-type tailspike as 7 minutes at 30ºC, 60 minutes at 20ºC, and 10 hours at 10ºC (Danner et al., 1993; Danner and Seckler, 1993). Given that the previously published experiments were performed without the initial ice step and were performed using different concentrations of proteins, the fact that the published values and those estimated here are on the same order of magnitude for all three temperatures indicates that the refolding of wild-type tailspike chains from inclusion bodies mimics the refolding of natively purified tailspike protein.

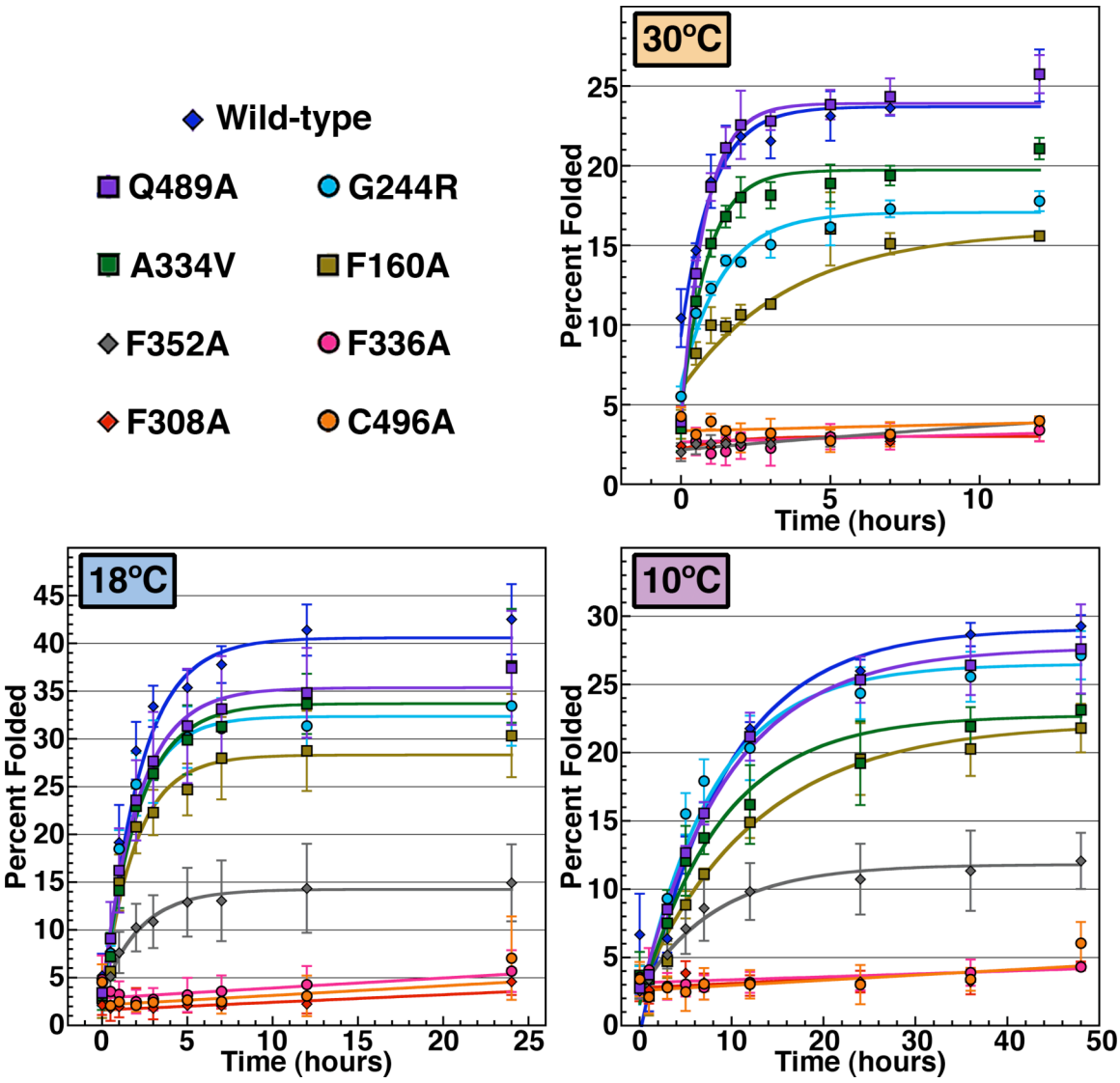

Figure 5-3. Kinetics of inclusion body refolding. For each time point, the percent of tailspike chains folded into SDS-resistant trimers was calculated. Colored symbols, as detailed in the legend in the upper left, represent averages of triplicate independent refolding experiments with vertical bars representing the standard deviations of these experiments. Colored curves are single exponential fits to the data, weighted by standard errors.

Table 5-1. Folding half-times of inclusion body refolding samples. The single exponential fits shown in Figure 5-3 were used to estimate the half-times of refolding. These values were calculated as the amount of time required for 50% of the fit's maximum yield to fold into a native, SDS-resistant trimer. Negative values, which are shaded and bolded, represent time courses that could not produce native tailspike.

Sample	*In Vitro* Trimer Refolding Half-Times (min)			
	37ºC	30ºC	18ºC	10ºC
WT	NA	12	85	256
F160A	NA	44	69	208
G244R	NA	20	71	212
F308A	NA	**-51**	**-245**	**-2220**
A334V	NA	24	76	215
F336A	NA	**-1269**	**-456**	**-3999**
F352A	NA	**-82**	58	144
Q489A	NA	24	75	219
C496A	NA	**-1763**	**-278**	**-1674**

Table 5-2. *In vitro* folding efficiencies of full length tailspike mutants. The percent of WT folding yields for 8 mutants and WT at the various refolding temperatures tested are given as averages and standard deviations of the final yields of multiple refolding experiments. Folding deficient average values (<75%) are shaded and bolded, while heavily defective values (<25%) are colored red.

Sample	*In Vitro* End-Point Percent of WT folding yields			
	37ºC	30ºC	18ºC	10ºC
WT	NA	100.0±6.4	100.0±8.6	100.0±2.7
F160A	NA	**60.8±0.2**	**71.4±10.3**	**74.5±6.1**
G244R	NA	**69.3±2.5**	78.7±9.8	92.7±6.0
F308A	NA	**13.5±3.0**	**10.8±3.2**	**14.9±1.1**
A334V	NA	82.1±2.6	88.6±14.0	79.0±4.0
F336A	NA	**13.3±2.8**	**13.4±5.1**	**14.7±0.9**
F352A	NA	**15.3±0.6**	**35.1±9.5**	**41.2±7.0**
Q489A	NA	100.4±4.7	88.0±14.0	94.3±11.2
C496A	NA	**15.5±1.1**	**16.6±10.3**	**20.6±5.3**

A number of mutants were able to refold to a native, trimeric state with kinetics similar to the wild-type full length tailspike protein. A334V, G244R, Q489A, and F160A were capable of producing detectable native trimer bands at all temperatures tested. As can be observed in Figure 5-3, although the final yields of these four mutants and wild-type tailspike were often significantly different, the time courses of these mutants tended to cluster with that of wild-type, as opposed to the time courses of F352A, C496A, F308A, and F336A. At 30ºC, the half-times of mutants A334V, G244R, and Q489A were approximately twice that of wild-type, while F160A had a refolding half time nearly 4 times that of wild-type. Though this difference may be significant, these values are still on the same order of magnitude. When refolded at 18ºC and 10ºC, these four mutants fold with half-times nearly identical to that of the wild-type sequence.

The single alanine mutant F352A displayed significantly lowered folding yields at 18ºC and 10ºC, and was completely blocked from folding at 30ºC (Figure 5-3). The fact that this mutation is capable of refolding back to its native conformation at 18ºC and 10ºC, but not at 30ºC, mirrors its *in vivo* folding phenotypes at these temperatures (Table 4-3). It is unclear why the folding yields for F352A are so much lower than the other refolding samples. When the half-time of refolding was calculated for F352A at 30ºC, a negative value was obtained (Table 5-1). This proved to be the case for all the other mutants incapable of producing native tailspike and is likely an indicator of the inability of these chains to fold to the native conformation. Positive values were obtained for F352A at 18ºC and 10ºC, and these values were higher than, but still on the same order, as those observed for wild-type and the other mutants capable of refolding.

4. A Comparison of *In Vivo* and *In Vitro* Data

A simple comparison of the temperature dependence of the *in vivo* folding of tailspike sequences (Table 4-3) and their ability to refold *in vitro* (Table 5-1) demonstrates that *in vitro* refolding data recapitulates the *in vivo* folding phenotypes, as previously highlighted with F352A. Some differences can be observed, however, when the end point folding yields of *in vitro* refolding experiments (Table 5-2) are compared with *in vivo* folding efficiencies (Figure 5-4).

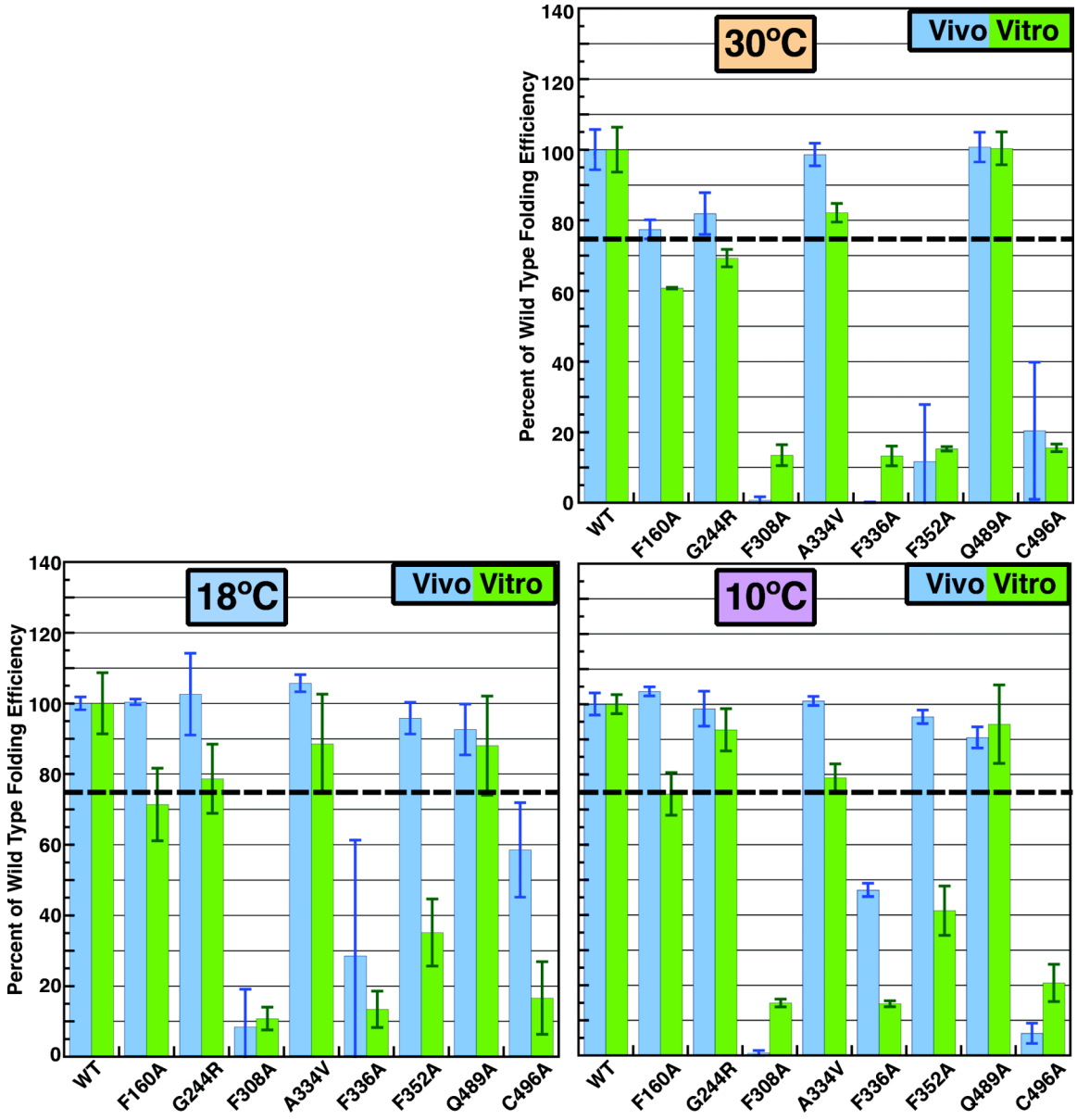

Figure 5-4. Quantitative comparison of *in vivo* and *in vitro* tailspike folding. Average percent of wild-type folding efficiencies are shown for the *in vivo* expression (blue) and *in vitro* refolding (green) of full length tailspike mutants at the three temperatures tested. Darker color lines indicate the standard deviations of triplicate experiments. The 75% threshold delineating wild-type-like folders and folding deficient samples is shown as a dashed line.

In comparing the *in vitro* and *in vivo* percent of wild-type folding efficiencies, a number of mutants displayed folding deficient values when refolded from inclusion bodies that did not display these values when folded in the environment of the cell. These inconsistencies, colored pink in Table 5-2, highlight F160A at all temperatures, F352A at lower temperatures, and G244R at 30ºC. Although these mutant samples would not have been classified as wild-type-like folders in the *in vivo* assay performed in Chapter 2, they do produce significant amounts of native tailspike, as is observed in Figure 5-3. Therefore, the ability to form native tailspike is consistent from *in vivo* expressions to *in vitro* refoldings.

Nonetheless, large differences between *in vivo* and *in vitro* conditions were observed for F352A at 18ºC and 10ºC, C496A at 18ºC, and F336A at 10ºC (Figure 5-4). The reason for these differences is unclear, but it is important to note that C496A, for example, is able to generate some SDS-resistant trimers *in vivo*, but was completely incapable of doing so *in vitro*. Many factors, such as the oxidation state of the refolding environment or the presence of other proteins in the inclusion body samples may affect the folding of these chains.

D. DISCUSSION

1. Single Alanine Mutations Block Refolding

Although some differences are apparent between the folding of full length mutant chains *in vivo* and *in vitro*, the temperature sensitivity displayed by mutant chains refolding from chemically denatured inclusion body pellets is consistent with their ability to fold *de novo* in the cell. In particular, F352A was capable of refolding at 18ºC and 10ºC, but was not able to refold at 30ºC. This mutant displays the same temperature sensitive folding phenotype when expressed in *E. coli*. Given that the refolding environment tested here is substantially different from the environment a nascent chain experiences inside the cell, the observed consistency of folding phenotypes supports the notion that the inability of these mutant chains to fold is due solely to the introduced alanine mutation.

Not only is it apparent that the single alanine mutations affect the folding and refolding of tailspike chains in a temperature sensitive manner, it is surprising that some mutations, such as F308A, are able to completely block the folding of the tailspike protein under all conditions so far tested. In the case of F308A, the deletion of a single side chain from a 666 amino acid polypeptide chain completely prevents the protein from folding in the cell, completely blocks inclusion body purified polypeptide chains from folding in the test tube, and bars βHX proteins from folding *in vivo*.

2. Sequence, not Ribosomal Interactions, Control Folding of β-helix

Mutants capable of folding upon expression *in vivo* were also able to fold into trimeric tailspike proteins upon renaturation from denatured inclusion bodies. This demonstrates that the observed folding phenotypes are not a product of cotranslational folding or the environment that the nascent chain experiences upon emergence from the ribosome exit tunnel. Instead, this is a product of the sequence itself.

The inclusion body samples were only approximately 70 – 75 % pure. Although this purity is consistent with a previously published purity of 68% for inclusion bodies generated with β-lactamase at 42ºC (Valax and Georgiou, 1993), it is clear that contaminants are present in the samples. It is possible that chaperones or ribosomal components may exist amongst those contaminants, as inclusion bodies have been observed to contain the heat-shock chaperone DnaK and ribosomal RNAs (Hartley and Kane, 1988; Rinas et al., 2007). These contaminants may assist or hinder the refolding of tailspike chains in a sequence specific manner. The fact that the observed kinetics of refolding of tailspike chains from inclusion bodies are consistent with the published values for the natively purified tailspike protein indicates that inclusion body contaminants do not significantly affect the rate of refolding. The unexpectedly low folding yields for all sequences tested may indicate that these contaminants do affect the relative rates of aggregation. This has been previously noted in the case of the refolding of hen egg white lysozyme from inclusion bodies, where contaminants affected the rate of aggregation, and hence folding yields, but did not disturb the rate of folding (Maachupalli-Reddy et al., 1997).

CHAPTER SIX:

IN VITRO REFOLDING OF β-HELIX MUTANT IN THE ISOLATED β-HELIX DOMAIN

A. INTRODUCTION

1. Fluorescence Studies on the Isolated β-helix Domain

Fluorescence intensity experiments have been utilized to characterize the reversible folding of the isolated β-helix domain and the affects that mutations have on this process (Miller et al., 1998a; Schuler and Seckler, 1998). An apparently two-state fluorescence transition has been demonstrated for the urea induced equilibrium unfolding and refolding of the β-helix domain that is similar to the unfolding and refolding transitions of the trimeric ΔN tailspike protein. The fluorescence kinetics of refolding are identical to that of ΔN tailspike, indicating that the isolated β-helix domain adopts a tertiary structure that resembles that of the β-helix domain in the full length protein.

Although the reversible folding of the isolated β-helix domain has allowed for the calculation of differences in free energies of folding for the wild-type sequence and various mutations, these experiments were performed under conditions that suppressed aggregation. When equilibrium studies were performed under strong ionic conditions, such as in the presence of guanidinium or urea with added sodium chloride, the fluorescence transition was no longer two-state but showed a plateau region in the equilibrium curve at 1 M guanidine or between 2 M and 4 M urea in the presence of 0.5 M salt (Schuler et al., 1999). This plateau matched a similar plateau observed in the refolding fluorescence of the full length tailspike protein (Fuchs et al., 1991). In the urea induced unfolding and refolding experiments, the transition region at 4 M urea from this plateau to the unfolded state was congruent with an apparently two-state acquisition of secondary structure, as measured by circular dichroism.

When high ionic strength equilibrium unfolding and refolding samples were allowed to incubate for extended times at 10ºC, it became apparent that conformations associated with the 2 M fluorescence transition and the plateau region were highly aggregation prone. The resulting aggregated protein exhibited multiple morphologies, including an amyloid-like fibril. When the affect of mutations was investigated on these aggregation processes, Seckler and coworkers observed that the temperature sensitive

folding mutant G244R substantially deviated from two-state fluorescence behavior at 18ºC, yielding a fluorescence intensity 33% lower than wild-type at low urea concentrations. Under these same conditions, the suppressor V331G displayed near two state behavior, but showed a dip in fluorescence intensity at 1.5 M urea. These data suggest that the β-helix domain folds through an aggregation prone intermediate that is more readily populated by the temperature sensitive folding mutant G244R.

2. Experimental Design

In Chapter 5, refolding experiments using purified inclusion bodies demonstrated that the buried core alanine mutants affected the refolding of the full length tailspike protein in a sequence specific and temperature sensitive manner that mirrored the *in vivo* folding phenotypes characterized in Chapter 2. This observation, in combination with the *in vivo* folding phenotypes of isolated β-helix domain mutants described in Chapter 4, supports the notion that single amino acid changes block the folding of the entire trimeric 210 kDa protein by altering the necessary information for encoding the folding of the β-helix domain. Can this effect be more directly demonstrated if isolated β-helix domain mutants are purified and refolded *in vitro*?

The answer to this question was investigated directly by purifying the wild-type isolated β-helix domain and the single alanine mutant F308A-βHX. Because the F308A mutation has proven to be completely blocked under all expression conditions tested, both constructs were purified under denaturing conditions and subsequently assayed for their ability to refold by fluorescence spectroscopy.

B. MATERIALS AND METHODS

1. Production and Purification of Isolated β-helix Constructs

Wild-type and F308A mutant pET(*βHX*) constructs were previously generated and transformed into *E. coli* BL21(DE3) as described in Chapter 4. Freeze stocks of βHX were streaked out on LB plates containing 100 μg/ml ampicillin and grown overnight at 37ºC. Clones from these plates were used to inoculate 37ºC overnight growth cultures in LB containing 100 μg/ml ampicillin. Expression cultures were inoculated via 200-fold dilution of the overnight cultures into 1 liter LB media containing 100 μg/ml ampicillin. Cells were grown shaking (225 rpm) for approximately 2.5 hours at 37ºC until cultures reached an OD_{600} between 0.6 and 0.7. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. Expression cultures were allowed to express protein for 7.5 hours at 37ºC. Cultures were harvested via centrifugation for 15 minutes at 4000*g*. Cell pellets were frozen and stored at -80ºC.

Frozen expression culture pellets were thawed at room temperature and lysed by mixture with 5 ml of denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris base, 8 M urea, pH 8.0) per gram of cell pellet for one hour at room temperature. Cellular debris was removed by centrifugation of lysates for 30 min at 10,000*g*. The resulting supernatant was mixed with Ni-NTA resin (Qiagen) and purified by gravity flow affinity chromatography. The Ni-NTA and lysate column was washed with denaturing lysis buffer and subsequently washed with denaturing wash buffer (100 mM NaH_2PO_4, 10 mM Tris base, 8 M urea, pH 6.3). His-tagged proteins were eluted with denaturing elution buffer (100 mM NaH_2PO_4, 10 mM Tris base, 8 M urea, pH 4.5). Elution fractions were pooled and concentrated using Amicon Ultra-15 centrifugal filters (Millipore).

Protein concentrations were determined by measuring absorbance at 280 nm of unfolded proteins, according to the Edelhoch method (Edelhoch, 1967). Extinction coefficients, 1.1872 $(mg/ml)^{-1}cm^{-1}$ for WT-βHX and 1.1891 $(mg/ml)^{-1}cm^{-1}$ for F308A-βHX, were calculated based on the protein sequence (Gill and von Hippel, 1989).

2. Equilibrium Refolding of the Isolated β-helix

βHX proteins were unfolded in 6 M urea (pH 7), 45 mM sodium phosphate, 0.9 mM EDTA, and 4.5 mM DTT at a concentration of 100 μg/ml for one hour at room temperature. These unfolding reactions were then diluted 10-fold into varying concentrations of urea solution (0 M – 6.5 M), giving a lowest final urea concentration of 0.6 M. Refolding samples were allowed to equilibrate for up to 5 days at 10ºC. The exact urea concentration of each sample was determined by measuring the refractive indexes of samples and applying the following equation:

$$[\text{Urea}] = 117.66*\Delta N + 29.753*\Delta N^2 + 185.56*\Delta N^3$$

where ΔN is the refractive index of the sample minus the refractive index of buffer lacking any urea.

3. Fluorescence Spectroscopy

Fluorescence emission spectra were taken using a Hitachi F-4500 fluorimeter equipped with a temperature control circulating water bath to maintain a constant temperature of 10ºC. The fluorimeter parameters included a slit width of 10 nm for excitation monochromators and 5 nm for emission monochromators. Samples were excited at 280 nm and the emission fluorescence was recorded for the wavelength range of 300 – 400 nm. All spectra were corrected for buffer.

C. RESULTS

1. Fluorescence of the Unfolded and Refolded States of βHX

Both wild-type βHX and F308A-βHX were purified by metal affinity chromatography to >90% homogeneity. Purified proteins were chemically denatured at neutral pH and refolded into varying concentrations of urea. To acquire the data shown here, DTT was required in the unfolding buffer. The range of urea that βHX chains experienced was 0.6 M to 6.4 M. Refolding reactions were allowed to equilibrate via incubation at 10ºC for 5 days. Equilibrium samples were then assayed for folding by fluorescence spectroscopy.

Completely unfolded samples mixed with 6.4 M urea were excited at 280 nm and monitored for fluorescence across a range of wavelengths from 300 nm to 440 nm. Both WT-βHX and F308A-βHX samples displayed nearly overlapping curves, with peaks at 350 nm for both samples (Figure 6-1). The WT-βHX sample equilibrating at the lowest concentration of urea, 0.6 M, showed a peak shift to 340 nm and a large increase in fluorescence intensity. The observed peaks and intensity changes are consistent with previously published fluorescence data on the WT-βHX protein (Miller et al., 1998a). This intensity increase was not observed if the unfolding buffer lacked DTT.

The F308A-βHX sample equilibrated at 0.6 M urea displayed a peak shift to 340 nm, but did not display a large intensity change (Figure 6-1). This is indicative that the conformation of this protein at low denaturant concentrations is not in the same folding state as the natively folded WT-βHX protein, nor is it in the completely unfolded state observed at high urea concentrations. The F308A-βHX protein appears to go into a third conformational state at low urea concentrations where some partial burial of tryptophans likely account for the peak shift and slight increase in fluorescence intensity.

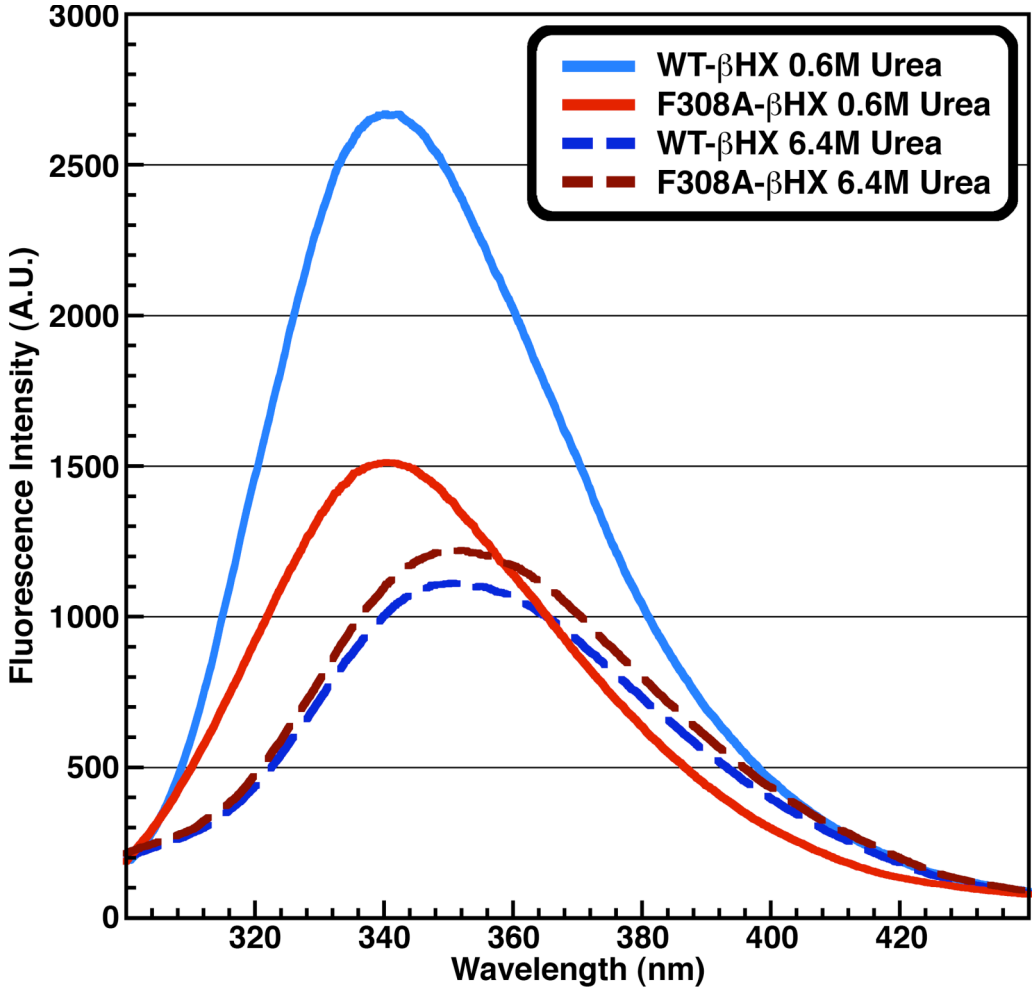

Figure 6-1. Unfolded and folded βHX fluorescence emission spectra. Fluorescence emission of wild-type βHX chains (blue) and F308A βHX chains (red) were monitored from 300 nm to 440 nm after excitation at 280 nm. Both samples were equilibrated at either 0.6 M urea (solid lines) or 6.4 M urea (dashed lines) for 5 days at 10ºC.

2. Equilibrium Refolding Fluorescence

The above fluorescence analysis and previously published data indicates that the fluorescence intensity of the β-helix domain at 340 nm is an indicator of the native folded state of the protein. Fluorescence emissions at this wavelength was observed for WT-βHX and F308A-βHX samples that had been refolding for five days in a range of urea concentrations

The WT-βHX protein displayed near two-state behavior, showing a clear transition between the unfolded state at high urea concentrations and the highly fluorescent native state at low urea concentrations. The midpoint of this transition occurred at approximately 2.5 M urea, which is in agreement with previously published data (Miller et al., 1998a; Schuler and Seckler, 1998). The only deviation from this two-state refolding was a dip in fluorescence intensity centered around 1.5 M urea. This is the concentration of urea at which refolding experiments that include high amounts of salt show a folding transition from an intermediate plateau to the native state (Schuler et al., 1999). This is also the concentration at which a similar dip in fluorescence can be observed in the 18ºC incubation of the suppressor mutation V331G. This dip in fluorescence may indicate a range of urea concentrations where the WT-βHX protein populates a partially folded, aggregation prone state. Although the aggregation of these samples was analyzed by absorbance at 360 nm, which measures light scattering due to high molecular weight species, turbidity measurements did not show any signal of aggregation. This is likely due to the fact that the 10 μg/ml concentration of proteins in the refolding samples was below the detection limit of the spectrophotometer.

F308A-βHX showed a clear deviation from two-state behavior (Figure 6-2). This deviation is nearly identical to that of G244R-βHX when it was incubated in high salt concentrations at 18ºC. The deviation is not only similar in the shape of the resulting equilibrium curve, but in the magnitude of the fluorescence signal as compared with the wild-type isolated β-helix domain. At the lowest tested urea concentration, the fluorescence of F308A-βHX was 44% less than that observed for WT-βHX.

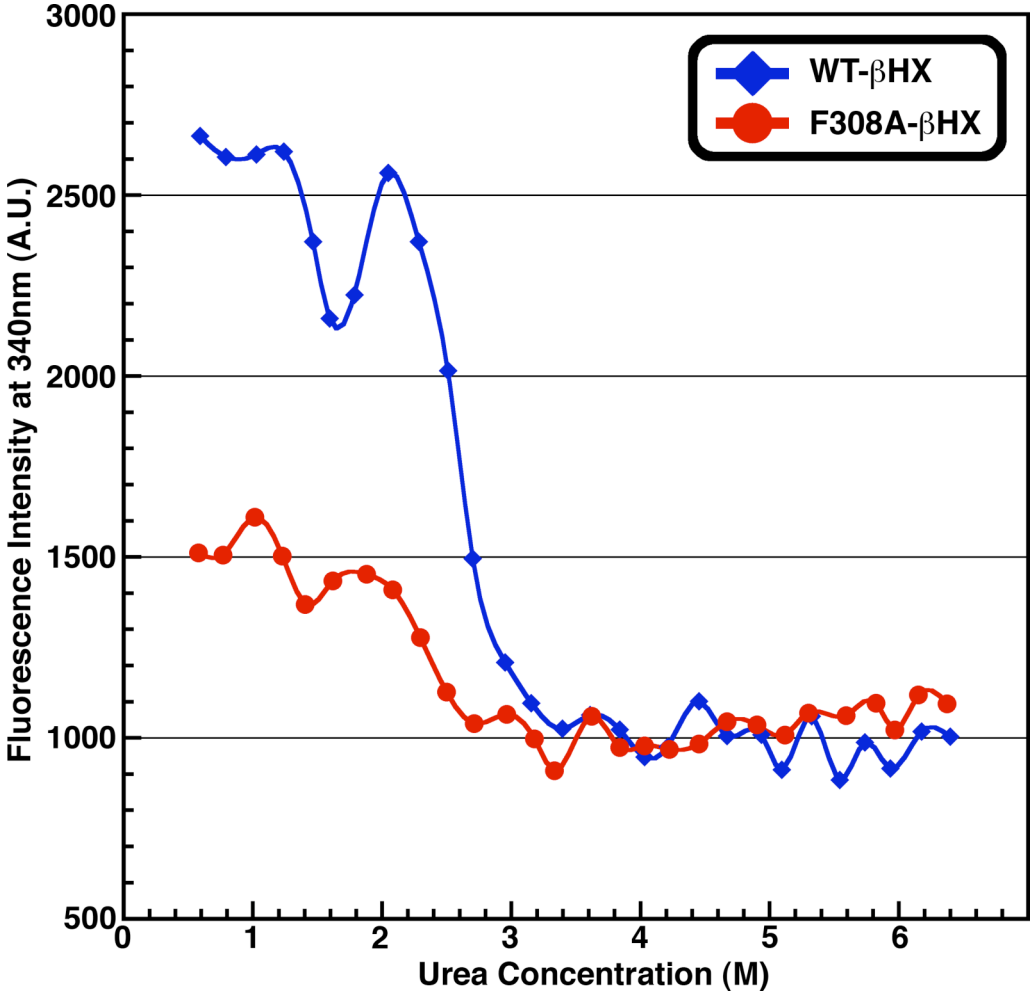

Figure 6-2. Equilibrium refolding of WT-βHX and F308A-βHX. Refolding samples containing 10 μg/ml protein were equilibrated for 5 days at 10ºC in various concentrations of urea. Fluorescence emission at 340 nm was examined after excitation at 280 nm. Solid dots represent the fluorescence intensity while lines are present simply as a guide.

D. DISCUSSION

1. Mutant β-helix Folding Phenotypes are due to Sequence Alone

WT-βHX and F38A-βHX protein were purified under denaturing conditions to a high degree of homogeneity. When assayed for their ability to refold, the WT-βHX folded to its native state, as assayed by fluorescence, while the F308A-βHX protein could not reach the same conformation. These experiments indicate that it is the polypeptide chain sequence of the single alanine mutation F308A that prevents the folding of the β-helix domain.

This single change appears to hinder the 42 kDa tailspike fragment from fully folding. Instead, the observed peak shift and partial fluorescence recovery indicate that the F308A-containing polypeptide chain goes into a partially folded conformation that is aggregation prone. The partial increase in fluorescence intensity and the peak shift observed at low urea concentrations indicate that some of the tryptophans of the β-helix protein may become partially buried, though not to the extent observed in the native structure. Though this burial may be due to aggregation, it is important to note that this altered burial is due to amino acid changes that do not necessarily contact any of the six tryptophans present on the surface of the β-helix domain (Figure 6-3).

Based upon the published CD spectra of the wild-type sequence under conditions of high salt, the β-helix appears capable of attaining its high degree of β-sheet content at urea concentrations much higher than the fluorescence transitions observed here (Schuler et al., 1999). Although it is unclear how the F308A mutation affects β-sheet formation, this data would argue that the F308A mutant populates a partially folded state that likely has β-sheet content but lacks proper tertiary structure. The structure of such a partially folded conformer may be a β-helical topology where strands are improperly registered or a nucleus for folding cannot form. In such a collapsed structure, tryptophans could be partially buried but not folded into a native structure, while exposed hydrophobic sequences would be particularly prone to aggregation.

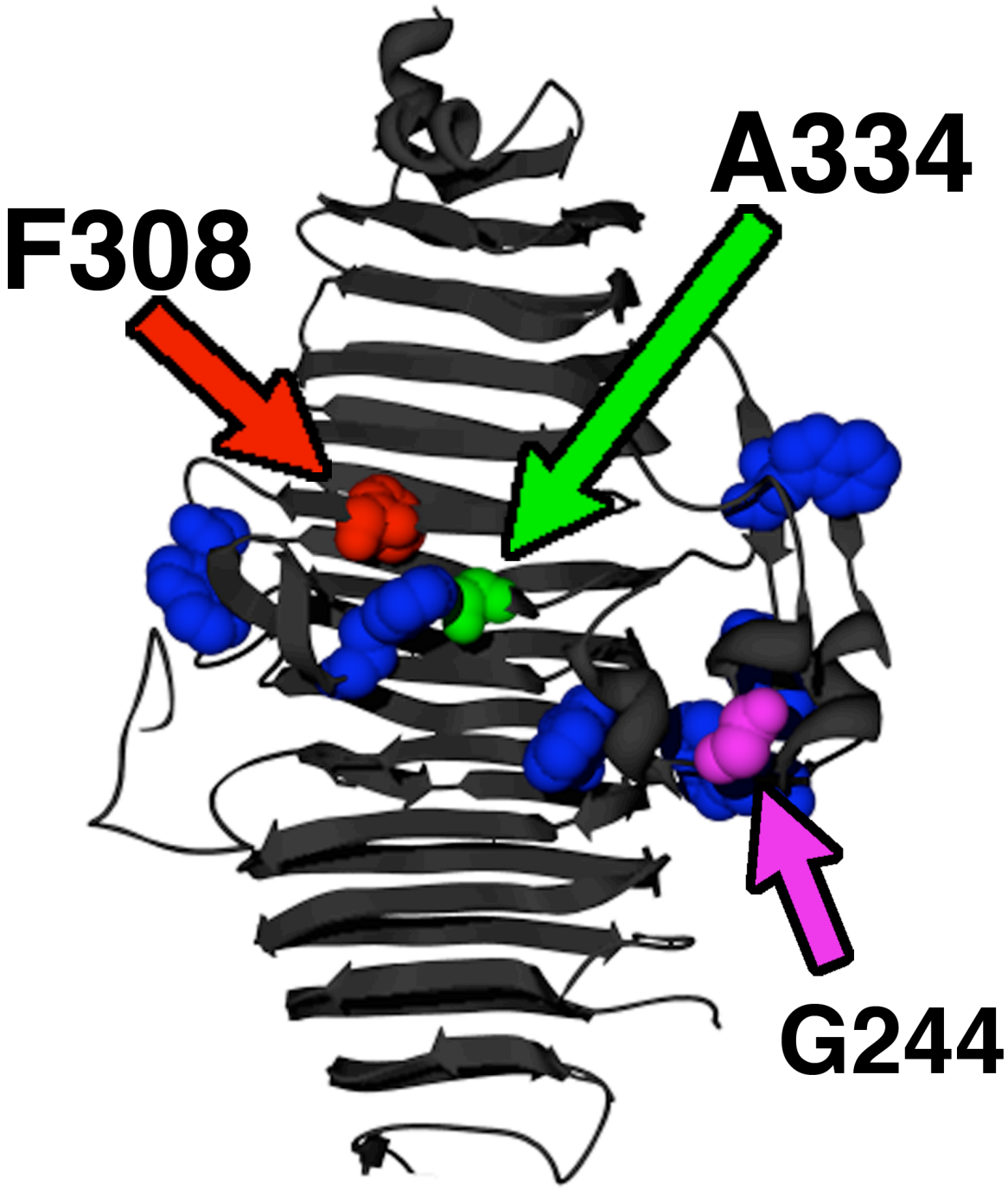

Figure 6-3. Locations of tryptophan residues in the tailspike β-helix domain. Six tryptophans, shown in blue, are located in the β-helix domain of the P22 tailspike. All are located on the surface of the protein. Three of these tryptophans (W202, W207, and W213) are located in the dorsal fin, while the other three (W315, W365, and W391) are on the surface of the β-helix. Residues A334 (green), F308 (red), and G244 (magenta) are shown for reference.

CHAPTER SEVEN:

DOUBLE MUTANT ANALYSIS REVEALS A NUCLEUS

A. INTRODUCTION

1. The Global Suppressors

Suppressor mutations of temperature sensitive folding mutations were repeatedly isolated in genetic screens at two positions in the P22 tailspike protein: V331 and A334 (Fane and King, 1991; Fane et al., 1991). Because these mutants intragenically rescued the temperature sensitivity of all mutants they were tested against, regardless of amino acid change or position, it was claimed that these suppressors acted in a global, or generic manner. Investigations of these mutants in the context of the isolated β-helix domain indicated that this global mechanism of suppression is likely via thermodynamic stabilization of a β-helical folding intermediate (Schuler and Seckler, 1998).

2. Experimental Design

The folding spine positions discovered in Chapter 2 represent the minimal required code for folding the β-helix domain. These mutations are incapable of folding like wild-type at all temperatures when expressed in the full length tailspike *in vivo*. For a small set of these mutants, it has been demonstrated that these alanine mutations block folding of the isolated β-helix domain (Chapter 4), block *in vitro* refolding of the full length protein (Chapter 5), and block the *in vitro* folding of the β-helix domain (Chapter 6). Can the global suppressor rescue these highly defective sequences?

If the global suppressor does rescue folding spine mutants, will the resulting folding efficiencies indicate which part of the β-helix domain folds first? After all, the central model of this thesis is that the β-helix domain of P22 tailspike folds via a nucleation-elongation mechanism. The folding spine does not appear to highlight a particular nucleus for folding. Does a specific region of the β-helix domain nucleate folding and, if so, where is it located?

To address these questions, double mutants of the single alanine folding spine mutations were made by introducing A334V into each single folding spine mutant sequence. The global suppressor A334V was chosen over other suppressor mutations

because the wild-type alanine side chain points into the core of the β-helix and because this mutant has been analyzed in a number of contexts relevant to this thesis (Baxa et al., 1999; Beissinger et al., 1995; Danner et al., 1993; Miller et al., 1998b; Mitraki et al., 1993; Mitraki et al., 1991; Schuler and Seckler, 1998). The C-terminal cysteine annulus mutations investigated in Chapter 2 were included in this investigation to control for the location of the mutants. Double mutants of G244R, ΔN, and ΔN-G244R were constructed to provide positive controls for rescue by A334V. These new double mutants were tested for their ability to fold *in vivo* at four temperatures: 37ºC, 30ºC, 18ºC, and 10ºC.

B. MATERIALS AND METHODS

1. Production of Double Mutants

Purified, methylated plasmids encoding confirmed single alanine mutants, the wild-type tailspike, the ΔN tailspike, and the mutated G244R tailspike were previously isolated from NovaBlue(DE3) cells (Novagen) or XL1-Blue cells as described in Chapter 2 and Chapter 5, Section B1. These plasmids were used as templates for introducing the A334V mutation via the PCR-based QuikChange mutagenesis procedure (Stratagene). The A334V mutagenic primer listed in Table 4-2 and its reverse complement were used to introduce this mutation in all but two of the single alanine mutant constructs. Since the codons encoding the G329A and F336A mutants are located in the region of the A334V mutagenic primer, new primers were designed to maintain these single alanine mutations while introducing the A334V mutation. The sequences of these primers (IDT) are listed in Table **7-1**. Mutagenesis reactions were performed in parallel in 96 well plates and utilized multi-channel pipeting to achieve high-throughput processing.

PCR products were transformed into *E. coli* BL21(DE3) cells as described in Chapter 2, Section B2. Plasmid DNA was purified using QIAprep Spin MiniPrep kits (QIAGEN). Purified plasmids were also submitted for DNA sequencing of a 400 base pair region (20% of the tailspike gene) containing the targeted codon. Sequencing confirmed the double mutant clones (MGH DNA Core, Cambridge, MA).

Table 7-1. Mutagenic primers used to construct double mutants. Forward mutagenic primers are given that were used to produce the specified mutations. Reverse primers used in the reaction were the exact reverse complement of the forward primers.

Mutation	Forward primer
G329A-A334V	5'-GCGTCAGTAAGTAGCGTCCAGTTTTTACGTAATAATGGTGGC-3'
F336A-A334V	5'-GGGTCAGTAAGTAGCGTCCAGGCTTTACGTAATAATGGTGGC-3'

2. Expression of Double Mutant Constructs

Clonal *E. coli* BL21(DE3) cultures of double mutants and the A334V single mutant were made into 40% glycerol working stocks and 8.3% DMSO freeze stocks after 8 hours of growth in LB containing 100 μg/ml ampicilllin at 37ºC. Working stocks were stored at -20ºC in 96 deep-well plates while freeze stocks were stored at -80ºC in dram vials. Overnight growth cultures were inoculated via 100-fold dilution of the working stocks into LB media containing 100 μg/ml ampicillin in preparation for inoculation of expression cultures. In the case of mutants with the F381A mutation, proper expression required that these cultures be freshly transformed into BL21(DE3) cells as described above and that the resulting transformant clones be used to inoculate overnight growth cultures.

Expression cultures were inoculated via 100-fold dilution of the overnight or the working stock cultures into LB media containing 0.5% glucose, 50 mM 3-[N-morpholino]-propanesulfonic acid (MOPS) and 100 μg/ml ampicillin. Cells were grown shaking (225 rpm) in 500 μl volumes in 96 deep-well plates for approximately 4.25 to 4.5 hours at 30ºC. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1.5 mM to induce protein expression. Cultures were transferred to 37ºC, 30ºC, 18ºC, or 10ºC and were allowed to express protein for approximately 4, 6.25, 18.3 hours, or over 30.5 hours respectively. Cultures were harvested via centrifugation for 15 minutes at 4000*g* and 4ºC. Cells were resuspended to 1/10 volume in lysis buffer (50 mM Tris-HCl, pH8.0; 25 mM NaCl; 2 mM EDTA; 0.1% Triton X-100) and stored at -20ºC.

3. Cell Lysis, Fractionation, and SDS-PAGE

Frozen expression samples were thawed at room temperature and lysed as described in Chapter 2, Section B3. Complete crude lysates, supernatant fractions, and pellet fractions were analyzed by SDS-PAGE as described in Chapter 2, Section B3.

4. Optical Densitometry

Optical densitometry was performed on coomassie stained gels and normalized folding efficiencies were calculated from complete crude lysate samples as described in Chapter 2, Section 4.

The amount of rescue conferred by the presence of A334V was calculated as the difference between the normalized folding efficiency of the A334V containing mutant and the normalized folding efficiency of the mutant lacking A334V. Standard deviations of this measure were calculated as the standard error of this difference:

$$\text{Standard Error of Difference} = \sqrt{\frac{(SD_{Mut})^2}{N_{Mut}} + \frac{(SD_{Mut+A334V})^2}{N_{Mut+A334V}}}$$

The above equation utilizes the standard deviation (SD) calculated for the single mutation (Mut) and its corresponding double mutant (Mut+A334V), as well as the number of independent expression samples (N) used in calculating the standard deviations.

C. RESULTS

1. *In Vivo* Expression of Double Mutants

All double mutants and single mutants not previously characterized, such as ΔN-G244R, were individually over-expressed from a pET plasmid in *Escherichia coli* cultures incubated at 37ºC, 30ºC, 18ºC and 10ºC in order to assess the affect of the A334V suppressor on previously characterized folding phenotypes. As was observed previously, all strains accumulated tailspike polypeptide chains as clearly resolved bands, indicating that the mutations did not recruit proteases. Host protein translational levels were consistent across samples, allowing for appropriate background subtraction so that the folding efficiencies of new expression cultures could be calculated accurately in the manner performed in Chapter 2. Folding efficiency values for strains not previously characterized were normalized to wild-type folding efficiencies. These newly calculated values are listed in Table 7-2.

Table 7-2. Folding efficiencies of new double mutants. The percent of wild-type folding efficiencies for double mutants at the four temperatures tested are given as averages and standard deviations of triplicate or sextuplicate independent expression experiments. Folding deficient average values (<75%) are shaded and bolded.

Sample	Percent of WT folding efficiency			
	37ºC	30ºC	18ºC	10ºC
ΔN-A334V	134.3±7.0	99.9±2.5	109.0±2.9	105.0±0.6
G244R-A334V	**64.6±32.1**	97.3±5.9	107.3±3.5	103.6±1.2
ΔN-G244R	**6.3±2.5**	**73.1±8.1**	107.3±3.4	105.0±1.5
ΔN-G244R-A334V	**73.9±22.0**	97.0±3.1	109.9±1.8	105.6±1.1
V141A-A334V	**8.7±7.0**	**16.6±7.7**	**62.5±9.0**	**52.9±7.8**
L144A-A334V	**0.7±0.5**	**1.3±1.1**	**28.7±8.8**	**29.7±13.4**
F152A-A334V	**6.6±4.9**	**0.0±0.0**	**11.5±9.1**	**13.8±3.9**
I167A-A334V	**0.2±0.2**	**0.0±0.0**	**32.6±10.7**	**46.8±15.4**
L179A-A334V	**0.1±0.1**	**0.0±0.0**	**21.9±9.0**	**49.6±16.1**
M195A-A334V	**0.1±0.1**	**0.1±0.0**	**43.7±17.8**	**72.9±10.9**
L262A-A334V	**0.3±0.4**	**0.5±0.5**	**16.8±5.0**	**35.2±14.6**
G277A-A334V	**0.3±0.4**	**0.2±0.2**	**7.2±1.4**	**20.6±6.5**
F282A-A334V	**1.0±0.9**	**0.3±0.1**	79.6±18.3	90.4±6.3
F284A-A334V	**0.1±0.1**	**0.0±0.0**	**42.2±8.8**	75.7±1.7
G300A-A334V	**0.4±0.4**	**1.2±0.3**	101.7±3.0	102.0±3.6
F308A-A334V	**0.0±0.0**	**0.0±0.0**	**5.8±3.1**	**25.2±9.7**
N310A-A334V	**0.2±0.2**	**0.0±0.0**	**67.2±8.7**	90.6±5.9
N319A-A334V	**0.1±0.1**	**0.0±0.1**	89.6±6.0	99.3±3.6
G324A-A334V	**0.1±0.1**	**0.8±1.0**	99.3±5.9	100.2±2.3
G329A-A334V	**0.4±0.5**	**8.5±3.1**	108.5±3.7	98.0±1.3
F336A-A334V	**0.3±0.4**	**0.5±0.5**	101.2±7.5	90.9±9.3
L379A-A334V	**0.2±0.2**	**0.0±0.0**	**2.4±2.2**	**3.8±2.1**
F381A-A334V	**0.0±0.0**	**0.0±0.0**	**0.0±0.0**	**0.1±0.0**
G393A-A334V	**0.6±0.7**	**0.1±0.0**	**0.8±0.9**	**2.3±1.1**
L396A-A334V	**0.1±0.1**	**0.0±0.0**	**0.2±0.2**	**1.3±1.0**

Sample	37°C	30°C	18°C	10°C
L424A-A334V	**0.4±0.3**	**3.9±1.7**	**16.4±3.6**	**31.8±12.9**
H426A-A334V	**0.1±0.1**	**0.8±1.1**	**1.2±1.1**	**1.7±1.6**
F441A-A334V	**0.6±0.6**	**0.1±0.1**	**0.4±0.5**	**0.0±0.0**
G445A-A334V	**0.8±0.6**	**0.0±0.0**	**0.5±0.6**	**0.0±0.0**
I453A-A334V	**0.5±0.7**	**0.4±0.5**	**0.0±0.1**	**0.0±0.0**
F472A-A334V	**1.0±0.7**	**0.1±0.1**	**1.2±1.0**	**1.5±1.0**
I475A-A334V	**0.6±0.6**	**0.3±0.3**	**0.6±0.5**	**0.3±0.3**
I492A-A334V	**0.8±0.6**	**0.7±0.5**	**0.8±0.3**	**2.0±1.2**
C496A-A334V	**0.4±0.2**	**0.3±0.4**	**0.1±0.1**	**0.0±0.1**
V537A-A334V	**49.5±33.7**	**63.3±9.6**	**60.1±9.3**	**54.9±17.0**
H610A-A334V	**0.2±0.3**	**0.2±0.2**	**0.0±0.0**	**0.0±0.0**
C613A-A334V	**0.4±0.2**	**0.2±0.1**	**0.0±0.0**	**0.1±0.1**
D632A-A334V	**1.3±0.6**	**0.1±0.1**	**0.2±0.2**	**0.4±0.3**
C635A-A334V	**1.9±0.7**	**2.7±1.1**	**3.7±2.1**	**4.6±2.0**
P638A-A334V	**2.6±0.7**	**1.9±1.0**	**9.2±0.2**	**29.4±5.4**

2. Mutant Folding Efficiencies at 37ºC and 30ºC

Folding spine mutants, as well as single alanine mutations in the cysteine annulus of the P22 tailspike, were previously shown to be deficient in folding at both 37ºC and 30ºC. This proved to be the case despite the presence of the A334V suppressor mutation (Figure 7-1). The suppressor mutant was completely incapable of rescuing any folding spine or cysteine annulus mutant back to wild-type-like folding efficiencies at 37ºC and 30ºC. For mutants that produced some amount of trimeric tailspike at these temperatures, the presence of the A334V suppressor did not significantly alter the folding of the mutant. This can be best seen with the mutation V537A, where the average folding efficiency of the single mutant was within a single standard deviation of its corresponding double mutant.

Although A334V did not significantly rescue any of the folding spine mutants, it did significantly hinder the folding of at least one of these mutants at 30ºC, V141A. As a single alanine mutant, this position displays a folding efficiency of 40±12%, but in conjuncation with the A334V mutation this mutant displays a folding efficiency of 17±8%. These values do not statistically overlap, demonstrating the unexpected result that the A334V suppressor might actually enhance folding deficient phenotypes, rather than cure them.

In contrast to the fact that A334V could not rescue the folding of any of the single alanine mutations, it was capable of significantly improving the folding of the G244R and ΔN-G244R control mutations at 37ºC and at 30ºC (Figure 7-1, left side). Although such rescue was expected at 37ºC, based on tailspike literature (Danner and Seckler, 1993; Fane et al., 1991; Mitraki et al., 1993), the fact that A334V could continue to enhance folding efficiencies of G244R mutants at 30ºC was an interesting result. A334V was also able to significantly improve the folding of the ΔN tailspike protein at 37ºC, but appeared to slightly hinder the folding of ΔN at 30ºC.

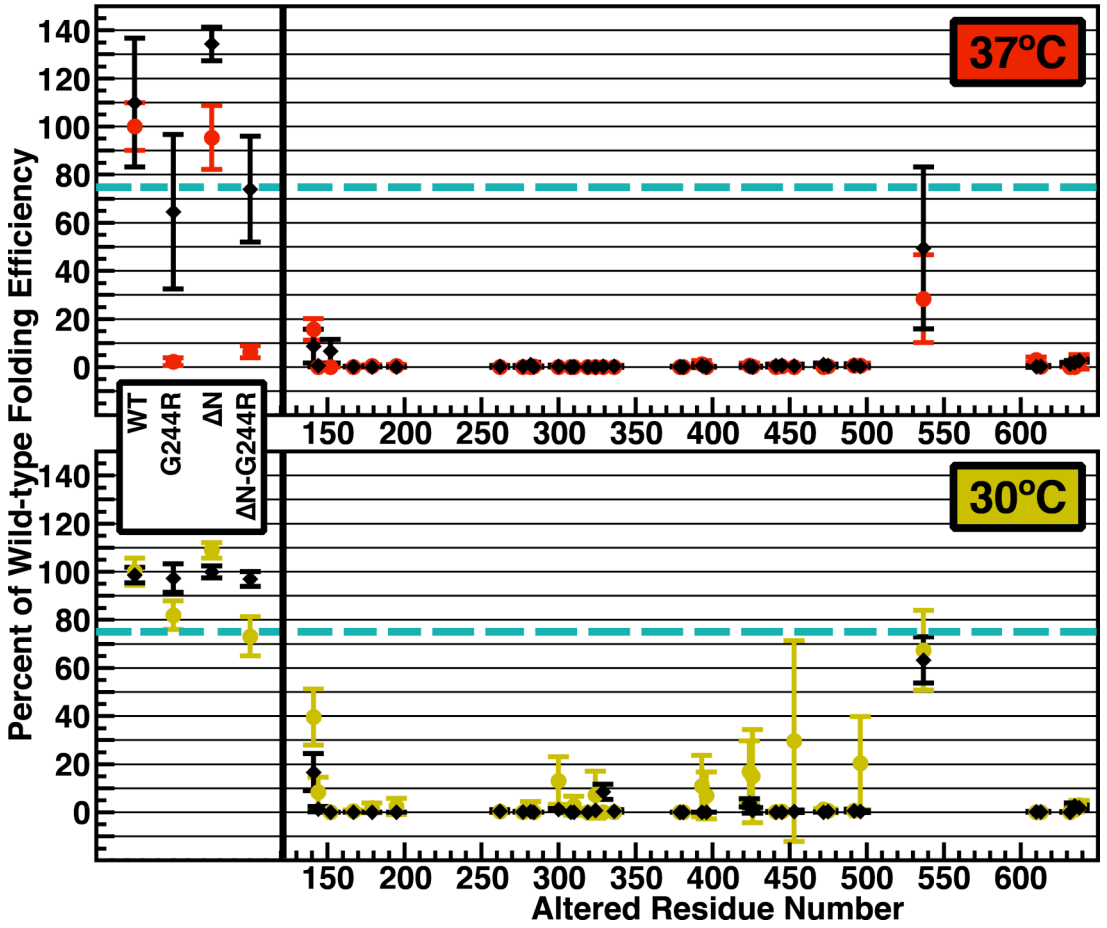

Figure 7-1. Folding efficiencies of single and double mutants at 37°C and 30°C. The average folding efficiencies of strains lacking the A334V suppressor mutation are represented by colored dots with their corresponding colored standard deviations. Average folding efficiencies and standard deviations of mutants containing A334V are shown as black diamonds and black bars respectively. Control strains are shown on the left of the vertical black bar, while alanine mutants are plotted on the right according to their position in the polypeptide sequence. A dashed cyan line is drawn at 75% folding efficiency.

3. Mutant Folding Efficiencies at 18ºC and 10ºC

At 18ºC and 10ºC, the presence of the A334V suppressor mutation did not significantly alter the folding efficiencies of the control mutations tested here, as all of these strains appeared to be at their maximal folding yields (Figure 7-2, left side).

In contrast, the presence of the suppressor mutation rescued a handful of folding spine mutations back to wild-type-like folding efficiencies (Figure 7-2, Table 7-2). The positions of the mutants rescued by A334V at 10ºC lie between residues number 282 and 336, and include the single alanine mutations F282A, F284A, G300A, N310A, N319A, G324A, G329A, and F336A. Of these mutants, only F284A and N310A were not rescued back to wild-type-like folding efficiencies at 18ºC, but were still significantly improved in their folding by the intragenic presence of A334V at this temperature.

Rescued alanine mutants are located close in sequence to the suppressor mutation. Some positions farther in sequence from the suppressor were also significantly improved in folding efficiency, though not enough to be classified as wild-type-like folders. These mutations include M195A at 10ºC, V537A at 18ºC and 10ºC, and P638A at 10ºC.

A number of mutations were significantly hindered in folding by the presence of the A334V mutant, particularly when expressed at 18ºC. These include I167A, L179A, G393A, L424A, I453A, I492A, C496A, and P638A.

4. A Region of Rescue

A334V rescues a handful of positions located close in sequence to itself at 18ºC and 10ºC. Although this can be seen in a general comparison of the folding efficiencies of single alanine mutants and their double mutant counterparts (Figure 7-2), the effects of the secondary-site mutation are most obvious when the exact percentage difference in folding efficiencies are calculated (Figure 7-3). For example, these differences highlight the rescue that A334V affords the G244R mutant in the full length and ΔN constructs. At 37ºC, A334V rescues G244R by 60 to 70%, while this rescue is only 15 to 25% at 30ºC. When these constructs are expressed at 18ºC and 10ºC, rescue drops to nil.

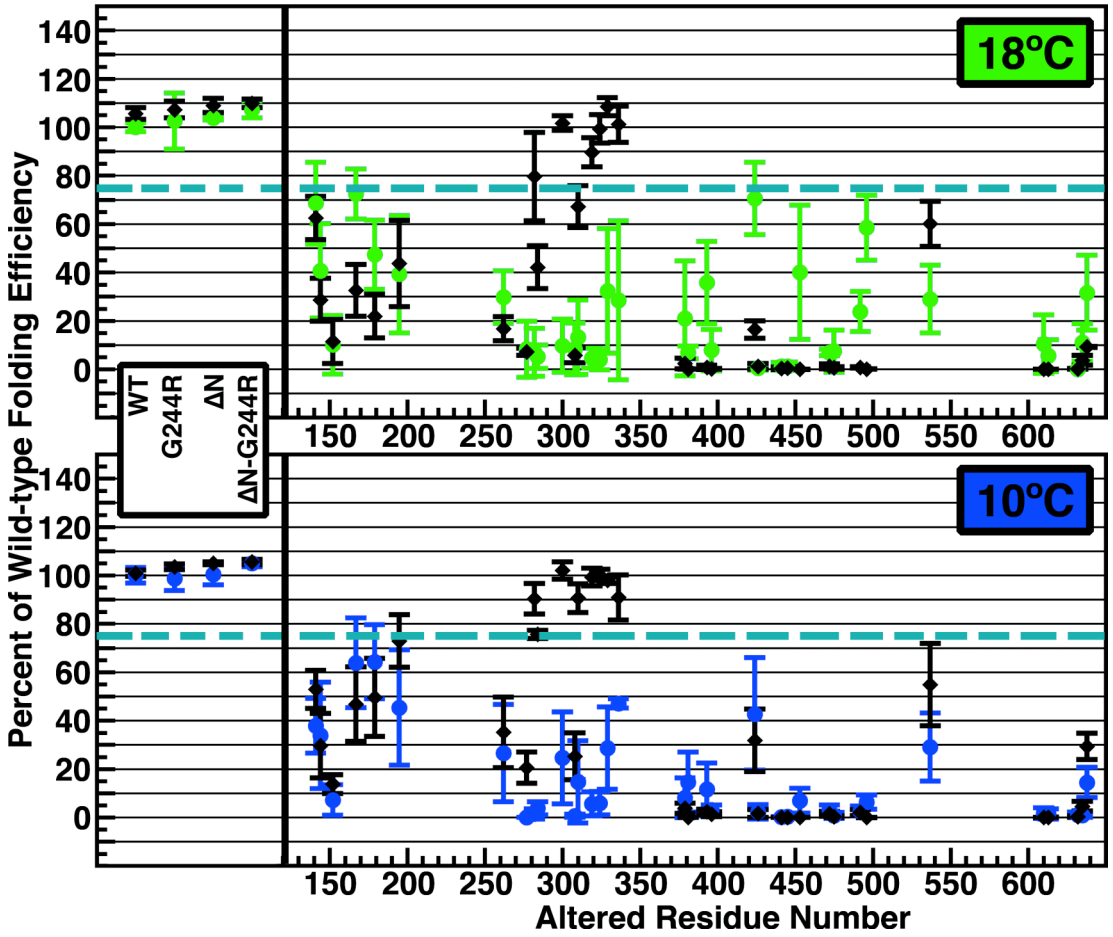

Figure 7-2. Folding efficiencies of single and double mutants at 18ºC and 10ºC. The average folding efficiencies of strains lacking the A334V suppressor mutation are represented by colored dots with their corresponding colored standard deviations. Average folding efficiencies and standard deviations of mutants containing A334V are shown as black diamonds and black bars respectively. Control strains are shown on the left of the vertical black bar, while alanine mutants are plotted on the right according to their position in the polypeptide sequence. A dashed cyan line is drawn at 75% folding efficiency.

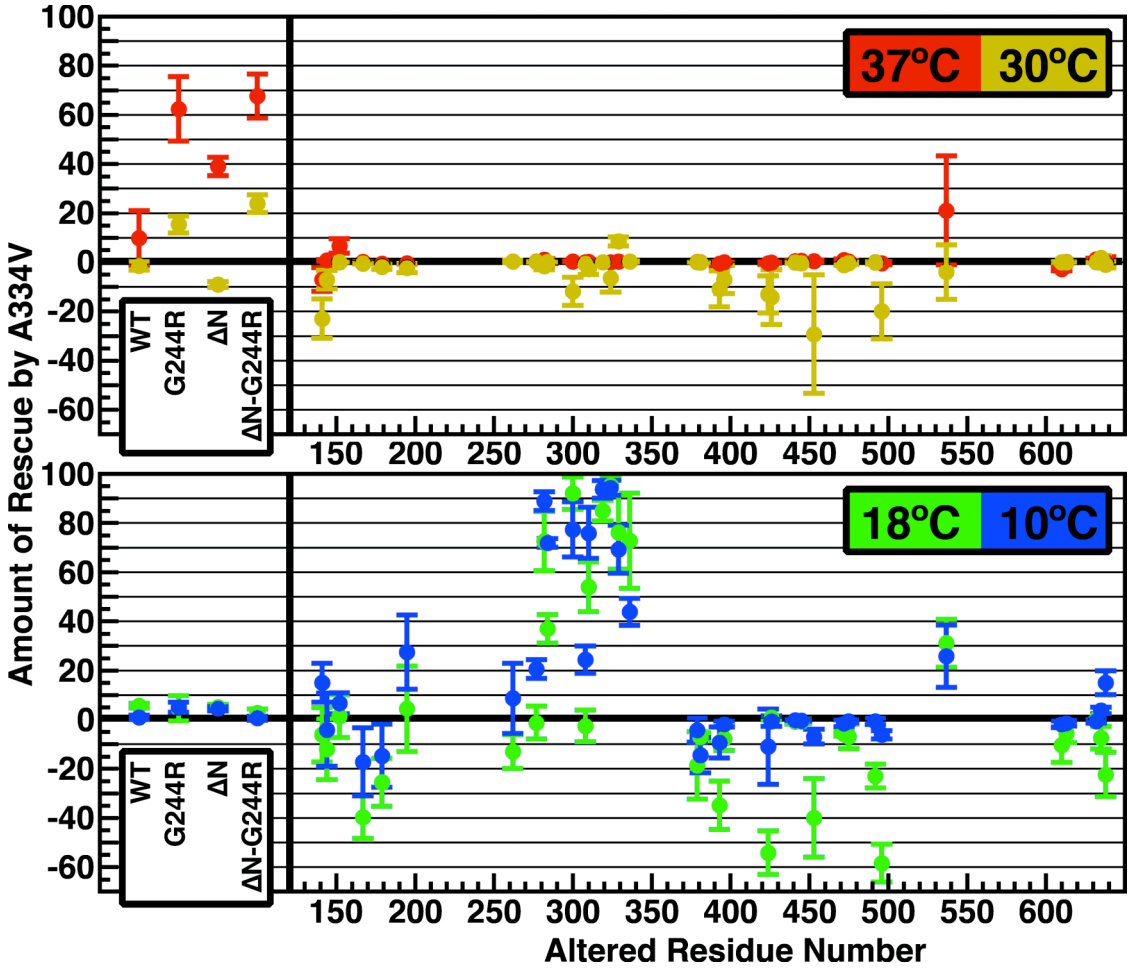

Figure 7-3. Quantitation of the amount of rescue afforded by A334V. Rescue was calculated by subtracting the folding efficiency of a base sequence from the folding efficiency observed for its corresponding A334V-containing double mutant. This value is plotted as a dot with error bars representing the standard error of differences between the two strains. Rescue of control sequences is shown on the left, while rescue of alanine mutants is plotted on the right according to their position in the polypeptide sequence.

For the single alanine mutants, 18ºC and 10ºC rescue values highlight the region just N-terminal to the position of the global suppressor (Figure 7-3). These positions not only include the mutants that became wild-type-like folders in the presence of A334V, but also include a number of other mutants that are significantly rescued by the suppressor, including M195A which displays a significant rescue value of 30% at 10ºC.

When the positions that are rescued by the A334V suppressor mutation back to wild-type-like levels, or alternatively are rescued by 50% or more at either 18ºC or 10ºC, are analyzed, the positions highlight a cluster of residues in rungs 4, 5 and 6 (Figure 7-4). Although these positions are in the general vicinity of the suppressor mutation, only one of these positions, F336, directly contacts the alanine at position 334 in the wild-type structure. Amongst the folding spine residues, the second closest position in space to alanine 334 is F308, whose alanine mutant cannot be rescued by the suppressor or by any temperature. Figure 7-4 also shows the positions of temperature sensitive mutants that were listed in the literature as being rescued by the global suppressor. These positions highlight the same region that the rescued folding spine positions highlight, as well as the dorsal fin. Folding mutants in this central portion of the β-helix appear to be rescueable by the A334V suppressor mutation.

5. A334V Hinders Folding of Distant Folding Spine Residues

Single alanine mutants outside the central rungs (4 – 6) of the β-helix have a greater tendency to be hindered by the presence of the suppressor mutation rather than be rescued by A334V (Figure 7-3). At 30ºC, 18ºC, and 10ºC, rescue values indicate that a number of single alanine mutations are significantly hindered by the presence of the suppressor mutation. These mutants, such as V141A at 30ºC, have negative rescue values that are significantly different from zero. The majority of these hindered positions are distantly C-terminal to the suppressor position. Many of these positions were hindered between 40 and 60% by the presence of the suppressor mutation. These are significant values, given that A334V rescues G244R by 60% at 37ºC.

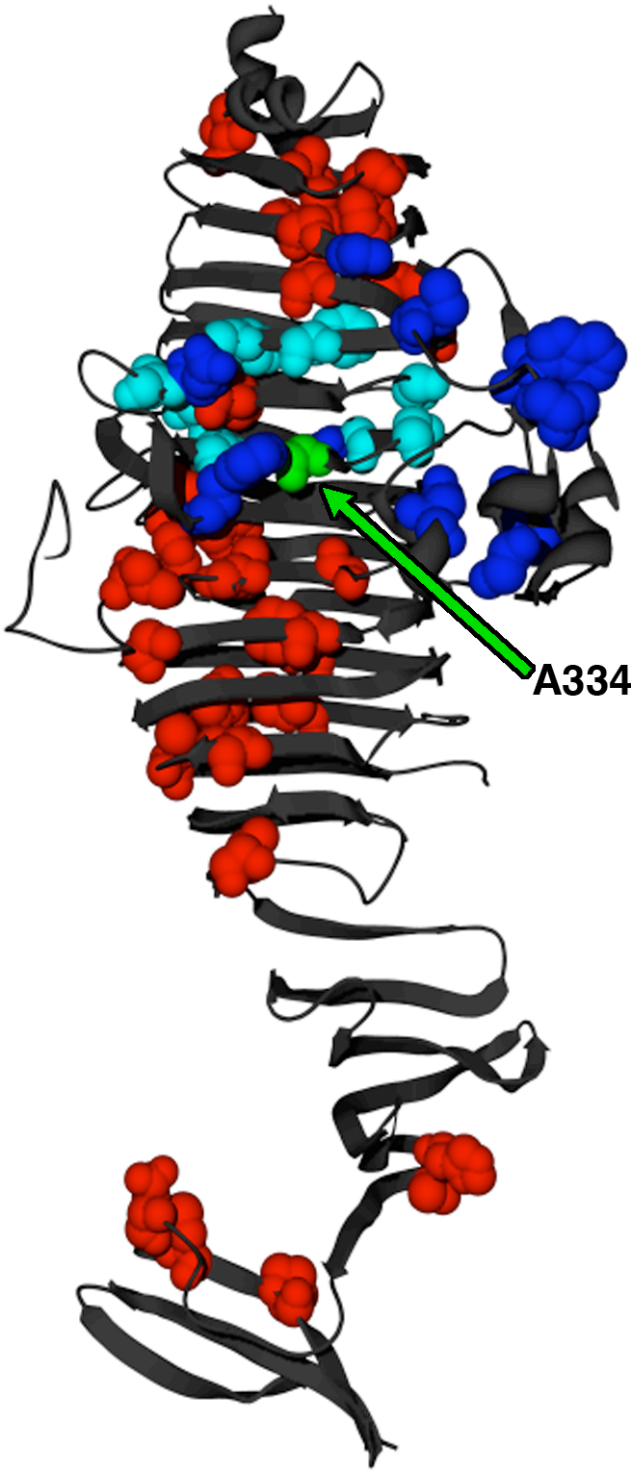

Figure 7-4. Locations of A334V rescued mutants. The native positions of folding spine mutants that are rescued back to wild-type-like folding efficiencies (>75%) by the presence of A334V (green) are shown in cyan, while the positions of other alanine mutants that are not rescued by A334V are colored red. Positions colored blue represent temperature sensitive mutations that have been previously published as being rescued by the suppressor mutation A334V.

D. DISCUSSION

1. A334V is Not a Global Suppressor

A334V and other suppressors of temperature sensitive folding defects were originally classified as global suppressors because they appeared to rescue the folding defects of all mutants tested via a general mechanism of suppression. Since the suppressor mutants thermodynamically stabilized the isolated β-helix domain, it was suggested that the generic mechanism of suppression was through a stabilization of the β-helical folding domain (Schuler and Seckler, 1998).

The results presented here demonstrate that the suppressor mutation A334V is not a global suppressor. This mutant is not capable of rescuing a number of alanine mutations located in the core of the β-helix that block the folding of the tailspike protein. In fact, A334V's pattern of suppression of folding spine mutations is position specific and is isolated to a specific region of the β-helix between rungs 4 and 6. This suggests that the mechanism by which A334V rescues folding defects is not by stabilizing the entire β-helix domain. If suppression occurred through enhanced thermostability of the β-helix domain, one would expect the suppressor to rescue positions in a manner more consistent with the type of mutations that are generated, rather than by location in the structure.

A number of folding defective alanine mutations are further hindered in folding by the presence of A334V. Since A334V has been shown to stabilize the β-helix, these results demonstrate non-additivity in the folding of the double mutant tailspike protein sequences. This precludes the possibility that the partitioning between productive folding and aggregation is dependent on the thermostability of the β-helical folding intermediate.

2. A Nucleus for Folding

If the suppressor mutation does not enhance productive protein folding through the thermodynamic stabilization of a β-helical folding intermediate, then perhaps it acts by stabilizing an earlier folding intermediate or a particular kinetic transition in the folding of the β-helix domain. Based on the nucleation-elongation mechanism proposed

in Chapter 2, the A334V mutant may be stabilizing a nucleus for folding or rescue only those mutations that disrupt the folding of the nucleus. The central region of the β-helix between rungs 4 and 6, where the suppressor mutation rescues folding spine mutations, most likely highlights this nucleus for folding.

If it is true that the A334V mutation acts by stabilizing a folding nucleus, then the non-additivity observed for mutations C-terminal to the suppressor is likely the result of the formation of a structure with a stable nucleus that is hindered in elongation by the folding spine alanine mutant. Such a combination of mutations would cause polypeptide chains to populate a conformation that is partially folded in the central part of the β-helix, likely exposes hydrophobic regions, and is highly prone to aggregation.

How does the suppressor mutation stabilize a nucleus for folding? Seckler has suggested that the substitution of valine for alanine at position 334 improves an internal hydrophobic stack, given that A334 is sandwiched between I306 and V362 in a buried core stack. The existence of the folding spine highlights the importance of key stacking positions for registration of one β-strand against the next in the β-helix topology. The A334V substitution may rescue folding spine alanine mutations in the nucleation region by stabilizing the proper registration of β-strands in the nucleus region. Alternatively, alanine mutations remove the necessary information for registering rungs against each other while the valine substitution at position 334 provides alternative information for β-sheet registration.

CHAPTER EIGHT:

FINAL DISCUSSION

A. SUMMARY OF CHAPTERS

Side chain stacking is a structural phenomenon that is a prominent component of protein structures, such as amyloid fibers and the β-helix fold, that are involved in a number of human illnesses. Neurodegenerative diseases such as Alzheimer's, Huntington's, and Parkinson's disease are coincident with the formation and accumulation of amyloid fibers at the sites of neuronal loss. Some allergies are the result of an immune response to the β-helix structure of antigens. The toxins responsible for whooping cough are β-helical in nature. At the end of Chapter 1, the question was posed as to how the amino acid sequence involved in side chain stacks directs the folding of these proteins.

This question was initially investigated in Chapter 1 by reviewing the known cases of side chain stacking. β-helix proteins, the amyloid crystal structure, and leucine rich repeat proteins all display varying degrees of side chain stacking. In the case of the amyloid crystal structure, side chain stacking involved identical residues stacking with identical orientations. In contrast, β-helix proteins tend to lack any form of sequence repeat between rungs, and β-helix stacks readily tolerate sequence identity mismatches. It was further observed that for the right-handed parallel β-helix structures, side chain stacking was more highly organized inside the buried cores of the proteins, while external side chain stacks were more structurally disorganized.

To address the question of how the amino acid sequence of buried core residues directs the folding of β-helix proteins, scanning alanine mutagenesis was performed on the buried core stacks of the P22 tailspike. In Chapter 2, alanine mutants of the buried core of the P22 tailspike were assayed for their ability to fold in the context of the cell. These experiments demonstrated that the buried core of the tailspike protein is surprisingly sensitive to alanine mutations at high expression temperatures. At lower temperatures, alanine mutations are tolerated more readily, though a core set of positions are completely intolerant to alanine substitution. These positions highlight a network of continuously contacting residues called the folding spine. The results of these experiments led to the proposal at the end of Chapter 2 that the β-helix domain of P22

tailspike folds through a step-wise nucleation-elongation mechanism, where a nucleus of a few rungs forms and acts as a template for subsequent rungs to wind up against. In this folding mechanism, the folding spine residues represent the necessary information for creating key contacts that dock one rung to the next. The remaining chapters of this thesis were dedicated to testing this model for the folding pathway of the P22 tailspike β-helix domain.

In Chapter 3, the native state stabilities of mutant tailspike proteins were assayed to determine if the alanine mutants investigated in Chapter 2 failed to accumulate native, trimeric tailspike proteins at various *in vivo* expression temperatures because the mutations destabilized the native state. A general thermostability assay demonstrated that all but one of the mutant proteins carrying alanine substitutions had melting temperatures between 80ºC and 90ºC in the presence of 2% SDS. Since these conditions are much harsher than that experienced by nascent chains folding in cells at the temperatures tested, the results established that the folding defects caused by single alanine mutations were not due to a destabilization of the native state.

To determine whether the buried core alanine mutants acted at the level of the folding of the β-helix domain of the full length tailspike protein, a small subset of mutants were investigated in the context of the isolated β-helix domain *in vivo*. The folding of the mutant βHX proteins was assayed by their solubility. The temperature sensitive folding phenotypes discovered for mutants in the isolated β-helix was consistent with those seen in the full length tailspike proteins. These experiments demonstrated that the inability of mutant chains to reach the native trimeric state in Chapter 2 was due to a defect in the folding of the β-helix domain.

Another important observation made in Chapter 4 is that multiple aggregation pathways exist for the P22 tailspike. This conclusion was borne out from the generation of ΔN-6H and βHX sequences. The ΔN-6H polypeptide chain was highly aggregative and was incapable of forming native, SDS-resistant trimers. βHX, which was identical in sequence to ΔN-6H except for the lack of the C-terminus of the tailspike protein, was able to form soluble, natively folded β-helix proteins. Therefore, in the case of ΔN-6H,

the C-terminus of the protein was responsible for aggregation while another aggregation pathway must be present for folding defective βHX mutants.

In Chapter 5, the *in vitro* refolding of full length tailspike mutants was assayed to determine if the sequence of these mutants alone is sufficient to cause the observed folding defects. Inclusion bodies were purified and denatured tailspike chains were investigated for their ability to refold to the native state. The competence of tailspike mutants to refold *in* vitro was consistent with *in vivo* folding phenotypes. The kinetics of refolding was also consistent with previously published half-times. Although *in vitro* folding yields were often reduced in comparison to the calculated *in vivo* folding efficiencies, the kinetics of refolding demonstrated that possible inclusion body contaminants did not have a significant effect on the folding of the tailspike. This indicates that the amino acid sequence of the tailspike polypeptide chains alone is sufficient to encode the folding of the protein. Further, the refolding experiments demonstrated that single alanine substitutions are sufficient for blocking the folding pathway.

In vitro refolding experiments were extended to the refolding of wild-type βHX and F308A-βHX proteins. These species were purified under denaturing conditions in Chapter 6 and equilibrium refolding experiments were monitored by fluorescence. The observed fluorescence signals indicated that the F308A mutation does in fact block the folding of this domain. However, the observed fluorescence values for F308A-βHX at low denaturant concentrations indicated that some burial of tryptophans occurs at low urea concentrations. This likely implies that the mutation F308A partially folds or collapses into an aggregation prone conformation.

The final test of the nucleation-elongation model was to ask whether the global suppressor mutation A334V could rescue folding spine mutations and demonstrate the existence of a folding nucleus. In Chapter 7, double mutants of the full length tailspike sequence were expressed *in vivo* and analyzed for the ability to form native trimers. Comparing the folding of single alanine mutations with the folding of double mutants containing A334V highlighted a region of residues in rungs 4, 5, and 6 that were rescued

by the suppressor mutation at 18ºC or 10ºC. I propose that this region represents a nucleus for the earliest of folding events in the β-helix domain.

Other folding spine residues could not be rescued by A334V and a number of mutants had significantly hindered folding efficiencies upon the introduction of the A334V mutant. The fact that A334V cannot rescue all mutations demonstrates that this mutant is not in fact a global suppressor. Although it has been previously demonstrated that A334V thermodynamically stabilizes the β-helix, the non-additivity of folding phenotypes of A334V and folding spine double mutants indicates that the thermodynamics of a native-like β-helix folding intermediate is not the primary controller of the folding of the β-helix domain. Instead, it appears that A334V stabilizes or kinetically enhances the population of a nucleus for β-helix folding.

B. MEDICAL IMPLICATIONS

As we come to better understand the basic principles behind misfolding and aggregation, we can begin to design therapeutics that aim to cure protein misfolding diseases. Already, a number of methods are utilized to combat these diseases. These methods fall into two categories: rational design and chemical library screening.

For example, a number of nonsteroidal anti-inflammatory drugs have been developed that specifically stabilize the native, tetrameric state of transthyretin (Miller et al., 2004). Stabilization of the native state prevents dissociation of the tetramer, which is a necessary step towards the progression of amyloid formation (Colon and Kelly, 1992). In the case of serpin polymerization, small peptides have been designed to bind into the middle strand position of β-sheet A, thereby blocking mobile loops of other proteins from intermolecular binding in that position and proceeding towards polymerization (Zhou et al., 2004). These two methods of therapeutics design are successful because they target native structures and prevent the disruption of that structure by aggregative processes.

The alternative discovery method has been applied to finding chemicals that can disaggregate amyloid fibrils associated with Alzhiemer's disease. This process resulted in

the discovery of 4,5-dianilinophthalimide (DAPH) which is capable of reversing Aβ(1-42) fiber formation, though the mechanism of action is currently unknown (Blanchard et al., 2004). Alternatively, this type of drug discovery process was used to find small chemicals that enhance inclusion formation (Bodner et al., 2006). These enhancers of aggregation alleviated some of the toxic effects of amyloidgenic Huntingtin protein and α-synuclein. It appears that forcing these proteins into inclusions prevents the protein from populating toxic species.

These methods of drug discovery and design have also been used to target pathogens, such as pathogenic gram-negative bactera. One example is the discovery of a pentadecapeptide that binds and inhibits the β-helical protein LpxA (Williams et al., 2006). LpxA catalyzes the first step in lipid A biosynthesis and is essential for growth. Inhibition of this protein therefore makes this pentadecapeptide a powerful antibiotic.

As can be seen from these examples of designed and discovered drugs, the main focus of modern therapeutics is to target the end states of the folding and aggregation processes. Drugs are either designed or screened for their ability bind the native state of a protein or to act on formed fibers. In the case of amyloidosis, this may not be ideal if the toxic species are small oligomeric species that would become more populated upon fiber disaggregation. Instead, a processive folding mechanism provides a new target for drug design against a number of diseases associated with β-helical proteins and protein relying on side chain stacking.

For example, it is reasonable that instead of blocking the interactions between an infective β-helix and its recognition factor, drugs may be designed to hinder the folding process. These may be particularly useful for combating disease-associated autotransporter β-helix proteins, whose folding and translocation are coupled (Otto et al., 2005). Inhibition of the folding process would prevent maturation of toxicity. It is clear that the core of the P22 tailspike β-helix is surprisingly sensitive to environmental factors and mutational effects, which can cause unrecoverable folding inhibition under physiological conditions. The same may be true for the folding of other structures that display side chain stacking. Therefore, side chain stacks and their associated folding

intermediates may represent an assortment of targets for inhibition of the step-wise folding or assembly of these proteins.

Since the β-helix class of proteins is rare and primarily specific to the microbial world, drugs of this type would benefit from specificity towards the infectious agent. Drugs capable of inhibiting β-helix formation may offer an effective mechanism for preventing selected microbial and viral infections.

CHAPTER NINE:

REFERENCES

Ali, M.H., Taylor, C.M., Grigoryan, G., Allen, K.N., Imperiali, B. and Keating, A.E. (2005) Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure. *Structure*, **13**, 225-234.

Aloy, P., Stark, A., Hadley, C. and Russell, R.B. (2003) Predictions without templates: new folds, secondary structure, and contacts in CASP5. *Proteins*, **53 Suppl 6**, 436-456.

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

Anfinsen, C.B. and Haber, E. (1961) Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem*, **236**, 1361-1363.

Anfinsen, C.B., Haber, E., Sela, M. and White, F.H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*, **47**, 1309-1314.

Anfinsen, C.B. and Scheraga, H.A. (1975) Experimental and theoretical aspects of protein folding. *Adv Protein Chem*, **29**, 205-300.

Baker, D., Sohl, J.L. and Agard, D.A. (1992) A protein-folding reaction under kinetic control. *Nature*, **356**, 263-265.

Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P.F., Mahadevan, D., Blundell, T.L. and Slingsby, C. (1990) X-ray analysis of beta B2-crystallin and evolution of oligomeric lens proteins. *Nature*, **347**, 776-780.

Baxa, U., Steinbacher, S., Weintraub, A., Huber, R. and Seckler, R. (1999) Mutations improving the folding of phage P22 tailspike protein affect its receptor binding activity. *J Mol Biol*, **293**, 693-701.

Beissinger, M., Lee, S.C., Steinbacher, S., Reinemer, P., Huber, R., Yu, M.H. and Seckler, R. (1995) Mutations that stabilize folding intermediates of phage P22 tailspike protein: folding in vivo and in vitro, stability, and structural context. *J Mol Biol*, **249**, 185-194.

Berget, P.B. and Poteete, A.R. (1980) Structure and functions of the bacteriophage P22 tail protein. *J Virol*, **34**, 234-243.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.

Betts, S., Haase-Pettingell, C., Cook, K. and King, J. (2004) Buried hydrophobic side-chains essential for the folding of the parallel beta-helix domains of the P22 tailspike. *Protein Sci*, **13**, 2291-2303.

Betts, S. and King, J. (1999) There's a right way and a wrong way: in vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Structure Fold Des*, **7**, R131-139.

Betts, S.D. and King, J. (1998) Cold rescue of the thermolabile tailspike intermediate at the junction between productive folding and off-pathway aggregation. *Protein Sci*, **7**, 1516-1523.

Bihoreau, M.T., Baudin, V., Marden, M., Lacaze, N., Bohn, B., Kister, J., Schaad, O., Dumoulin, A., Edelstein, S.J., Poyart, C. and et al. (1992) Steric and hydrophobic determinants of the solubilities of recombinant sickle cell hemoglobins. *Protein Sci*, **1**, 145-150.

Blanchard, B.J., Chen, A., Rozeboom, L.M., Stafford, K.A., Weigele, P. and Ingram, V.M. (2004) Efficient reversal of Alzheimer's disease fibril formation and elimination of neurotoxicity by a small molecule. *Proc Natl Acad Sci U S A*, **101**, 14326-14332.

Bodner, R.A., Housman, D.E. and Kazantsev, A.G. (2006) New directions for neurodegenerative disease therapy: using chemical compounds to boost the formation of mutant protein inclusions. *Cell Cycle*, **5**, 1477-1480.

Bonar, L., Cohen, A.S. and Skinner, M.M. (1969) Characterization of the amyloid fibril as a cross-beta protein. *Proc Soc Exp Biol Med*, **131**, 1373-1375.

Bradley, P., Cowen, L., Menke, M., King, J. and Berger, B. (2001) BETAWRAP: successful prediction of parallel beta -helices from primary sequence reveals an association with many microbial pathogens. *Proc Natl Acad Sci U S A*, **98**, 14819-14824.

Brunschier, R., Danner, M. and Seckler, R. (1993) Interactions of phage P22 tailspike protein with GroE molecular chaperones during refolding in vitro. *J Biol Chem*, **268**, 2767-2772.

Burley, S.K. and Petsko, G.A. (1988) Weakly polar interactions in proteins. *Adv Protein Chem*, **39**, 125-189.

Capaldi, A.P. and Radford, S.E. (1998) Kinetic studies of beta-sheet protein folding. *Curr Opin Struct Biol*, **8**, 86-92.

Carrell, R.W. and Gooptu, B. (1998) Conformational changes and disease--serpins, prions and Alzheimer's. *Curr Opin Struct Biol*, **8**, 799-809.

Caughey, B. and Lansbury, P.T. (2003) Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci*, **26**, 267-298.

Chang, J., Weigele, P., King, J., Chiu, W. and Jiang, W. (2006) Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure*, **14**, 1073-1082.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S. and Hultgren, S.J. (2002) Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science*, **295**, 851-855.

Chen, B. and King, J. (1991) Thermal unfolding pathway for the thermostable P22 tailspike endorhamnosidase. *Biochemistry*, **30**, 6260-6269.

Chen, J., Flaugh, S.L., Callis, P.R. and King, J. (2006) Mechanism of the highly efficient quenching of tryptophan fluorescence in human gammaD-crystallin. *Biochemistry*, **45**, 11552-11563.

Cheung, M.S., Klimov, D. and Thirumalai, D. (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci U S A*, **102**, 4753-4758.

Chiti, F. and Dobson, C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, **75**, 333-366.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*, **86**, 2172-2175.

Ciccarelli, F.D., Copley, R.R., Doerks, T., Russell, R.B. and Bork, P. (2002) CASH--a beta-helix domain widespread among carbohydrate-binding proteins. *Trends Biochem Sci*, **27**, 59-62.

Clantin, B., Hodak, H., Willery, E., Locht, C., Jacob-Dubuisson, F. and Villeret, V. (2004) The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proc Natl Acad Sci U S A*, **101**, 6194-6199.

Clark, P.L. (2004) Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci*, **29**, 527-534.

Clark, P.L. and King, J. (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem*, **276**, 25411-25420.

Clark, P.L., Liu, Z.P., Rizo, J. and Gierasch, L.M. (1997) Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. *Nat Struct Biol*, **4**, 883-886.

Clark, P.L., Liu, Z.P., Zhang, J. and Gierasch, L.M. (1996) Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci*, **5**, 1108-1117.

Clarke, D.T., Doig, A.J., Stapley, B.J. and Jones, G.R. (1999) The alpha-helix folds on the millisecond time scale. *Proc Natl Acad Sci U S A*, **96**, 7232-7237.

Colon, W. and Kelly, J.W. (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry*, **31**, 8654-8660.

Crowther, D.C., Serpell, L.C., Dafforn, T.R., Gooptu, B. and Lomas, D.A. (2003) Nucleation of alpha 1-antichymotrypsin polymerization. *Biochemistry*, **42**, 2355-2363.

Cunningham, B.C. and Wells, J.A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081-1085.

Czerwinski, E.W., Midoro-Horiuti, T., White, M.A., Brooks, E.G. and Goldblum, R.M. (2005) Crystal structure of Jun a 1, the major cedar pollen allergen from Juniperus ashei, reveals a parallel beta-helical core. *J Biol Chem*, **280**, 3740-3746.

Danek, B.L. and Robinson, A.S. (2003) Nonnative interactions between cysteines direct productive assembly of P22 tailspike protein. *Biophys J*, **85**, 3237-3247.

Danner, M., Fuchs, A., Miller, S. and Seckler, R. (1993) Folding and assembly of phage P22 tailspike endorhamnosidase lacking the N-terminal, head-binding domain. *Eur J Biochem*, **215**, 653-661.

Danner, M. and Seckler, R. (1993) Mechanism of phage P22 tailspike protein folding mutations. *Protein Sci*, **2**, 1869-1881.

Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci*, **24**, 329-332.

Dobson, C.M. (2006) Protein aggregation and its consequences for human disease. *Protein Pept Lett*, **13**, 219-227.

Doyle, S.M., Anderson, E., Parent, K.N. and Teschke, C.M. (2004) A concerted mechanism for the suppression of a folding defect through interactions with chaperones. *J Biol Chem*, **279**, 17473-17482.

Eanes, E.D. and Glenner, G.G. (1968) X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem*, **16**, 673-677.

Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948-1954.

Emsley, P., Charles, I.G., Fairweather, N.F. and Isaacs, N.W. (1996) Structure of Bordetella pertussis virulence factor P.69 pertactin. *Nature*, **381**, 90-92.

Fane, B. and King, J. (1991) Intragenic suppressors of folding defects in the P22 tailspike protein. *Genetics*, **127**, 263-277.

Fane, B., Villafane, R., Mitraki, A. and King, J. (1991) Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *J Biol Chem*, **266**, 11640-11648.

Ferrone, F.A., Hofrichter, J. and Eaton, W.A. (1985) Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. *J Mol Biol*, **183**, 611-631.

Fink, A.L. (2005) Natively unfolded proteins. *Curr Opin Struct Biol*, **15**, 35-41.

Flaugh, S.L., Kosinski-Collins, M.S. and King, J. (2005a) Contributions of hydrophobic domain interface interactions to the folding and stability of human gammaD-crystallin. *Protein Sci*, **14**, 569-581.

Flaugh, S.L., Kosinski-Collins, M.S. and King, J. (2005b) Interdomain side-chain interactions in human gammaD crystallin influencing folding and stability. *Protein Sci*, **14**, 2030-2043.

Flaugh, S.L., Mills, I.A. and King, J. (2006) Glutamine deamidation destabilizes human gammaD-crystallin and lowers the kinetic barrier to unfolding. *J Biol Chem*, **281**, 30782-30793.

Fomenko, D.E. and Gladyshev, V.N. (2003) Identity and functions of CxxC-derived motifs. *Biochemistry*, **42**, 11214-11225.

Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E. and Kelly, J.W. (2006) Functional amyloid formation within mammalian tissue. *PLoS Biol*, **4**, e6.

Fuchs, A., Seiderer, C. and Seckler, R. (1991) In vitro folding pathway of phage P22 tailspike protein. *Biochemistry*, **30**, 6598-6604.

Gage, M.J. and Robinson, A.S. (2003) C-terminal hydrophobic interactions play a critical role in oligomeric assembly of the P22 tailspike trimer. *Protein Sci*, **12**, 2732-2747.

Gage, M.J., Zak, J.L. and Robinson, A.S. (2005) Three amino acids that are critical to formation and stability of the P22 tailspike trimer. *Protein Sci*, **14**, 2333-2343.

Gao, Y., Vainberg, I.E., Chow, R.L. and Cowan, N.J. (1993) Two cofactors and cytoplasmic chaperonin are required for the folding of alpha- and beta-tubulin. *Mol Cell Biol*, **13**, 2478-2485.

Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, **182**, 319-326.

Goldenberg, D. and King, J. (1982) Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. *Proc Natl Acad Sci U S A*, **79**, 3403-3407.

Goldenberg, D.P., Berget, P.B. and King, J. (1982) Maturation of the tail spike endorhamnosidase of Salmonella phage P22. *J Biol Chem*, **257**, 7864-7871.

Goldenberg, D.P. and King, J. (1981) Temperature-sensitive mutants blocked in the folding or subunit of the bacteriophage P22 tail spike protein. II. Active mutant proteins matured at 30 degrees C. *J Mol Biol*, **145**, 633-651.

Gordon, C.L., Sather, S.K., Casjens, S. and King, J. (1994) Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. *J Biol Chem*, **269**, 27941-27951.

Graether, S.P., Kuiper, M.J., Gagne, S.M., Walker, V.K., Jia, Z., Sykes, B.D. and Davies, P.L. (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*, **406**, 325-328.

Gunasekaran, K., Eyles, S.J., Hagler, A.T. and Gierasch, L.M. (2001) Keeping it in the family: folding studies of related proteins. *Curr Opin Struct Biol*, **11**, 83-93.

Guo, Z. and Eisenberg, D. (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc Natl Acad Sci U S A*, **103**, 8042-8047.

Haase-Pettingell, C., Betts, S., Raso, S.W., Stuart, L., Robinson, A. and King, J. (2001) Role for cysteine residues in the in vivo folding and assembly of the phage P22 tailspike. *Protein Sci*, **10**, 397-410.

Haase-Pettingell, C. and King, J. (1997) Prevalence of temperature sensitive folding mutations in the parallel beta coil domain of the phage P22 tailspike endorhamnosidase. *J Mol Biol*, **267**, 88-102.

Haase-Pettingell, C.A. and King, J. (1988) Formation of aggregates from a thermolabile in vivo folding intermediate in P22 tailspike maturation. A model for inclusion body formation. *J Biol Chem*, **263**, 4977-4983.

Haber, E. and Anfinsen, C.B. (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem*, **236**, 422-424.

Hardcastle, A.J., Thiselton, D.L., Van Maldergem, L., Saha, B.K., Jay, M., Plant, C., Taylor, R., Bird, A.C. and Bhattacharya, S. (1999) Mutations in the RP2 gene cause disease in 10% of families with familial X-linked retinitis pigmentosa assessed in this study. *Am J Hum Genet*, **64**, 1210-1215.

Harper, J.D. and Lansbury, P.T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem*, **66**, 385-407.

Hartley, D.L. and Kane, J.F. (1988) Properties of inclusion bodies from recombinant Escherichia coli. *Biochem Soc Trans*, **16**, 101-102.

He, M.M., Wood, Z.A., Baase, W.A., Xiao, H. and Matthews, B.W. (2004) Alanine-scanning mutagenesis of the beta-sheet region of phage T4 lysozyme suggests that tertiary context has a dominant effect on beta-sheet formation. *Protein Sci*, **13**, 2716-2724.

Hegde, S.S., Vetting, M.W., Roderick, S.L., Mitchenall, L.A., Maxwell, A., Takiff, H.E. and Blanchard, J.S. (2005) A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. *Science*, **308**, 1480-1483.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. and Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev*, **68**, 692-744.

Huang, W., Matte, A., Li, Y., Kim, Y.S., Linhardt, R.J., Su, H. and Cygler, M. (1999) Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 A resolution. *J Mol Biol*, **294**, 1257-1269.

Huntington, J.A., Read, R.J. and Carrell, R.W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, **407**, 923-926.

Ignatova, Z., Krishnan, B., Bombardier, J.P., Marcelino, A.M., Hong, J. and Gierasch, L.M. (2007) From the test tube to the cell: exploring the folding and aggregation of a beta-clam protein. *Biopolymers*, **88**, 157-163.

Ingram, V.M. (1956) A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature*, **178**, 792-794.

Irving, J.A., Pike, R.N., Lesk, A.M. and Whisstock, J.C. (2000) Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res*, **10**, 1845-1864.

Israel, V. (1978) A model for the adsorption of phage P22 to Salmonella typhimurium. *J Gen Virol*, **40**, 669-673.

Iwashita, S. and Kanegasaki, S. (1976) Enzymic and molecular properties of base-plate parts of bacteriophage P22. *Eur J Biochem*, **65**, 87-94.

Janciauskiene, S., Eriksson, S., Callea, F., Mallya, M., Zhou, A., Seyama, K., Hata, S. and Lomas, D.A. (2004) Differential detection of PAS-positive inclusions formed by the Z, Siiyama, and Mmalton variants of alpha1-antitrypsin. *Hepatology*, **40**, 1203-1210.

Jenkins, J. and Pickersgill, R. (2001) The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol*, **77**, 111-175.

Jin, X., Ballicora, M.A., Preiss, J. and Geiger, J.H. (2005) Crystal structure of potato tuber ADP-glucose pyrophosphorylase. *Embo J*, **24**, 694-704.

Johnson, D. and Travis, J. (1978) Structural evidence for methionine at the reactive site of human alpha-1-proteinase inhibitor. *J Biol Chem*, **253**, 7142-7144.

Jung, W.S., Hong, C.K., Lee, S., Kim, C.S., Kim, S.J., Kim, S.I. and Rhee, S. (2007) Structural and functional insights into intramolecular fructosyl transfer by inulin fructotransferase. *J Biol Chem*, **282**, 8414-8423.

Junker, M., Schuster, C.C., McDonnell, A.V., Sorg, K.A., Finn, M.C., Berger, B. and Clark, P.L. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A*, **103**, 4918-4923.

Jurnak, F., Yoder, M.D., Pickersgill, R. and Jenkins, J. (1994) Parallel beta-domains: a new fold in protein structures. *Curr Opin Struct Biol*, **4**, 802-806.

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F. and Rossmann, M.G. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature*, **415**, 553-557.

Kay, M.S. and Baldwin, R.L. (1996) Packing interactions in the apomyglobin folding intermediate. *Nat Struct Biol*, **3**, 439-445.

Kim, J. and Robinson, A.S. (2006) Dissociation of intermolecular disulfide bonds in P22 tailspike protein intermediates in the presence of SDS. *Protein Sci*, **15**, 1791-1793.

Kim, P.S. and Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu Rev Biochem*, **59**, 631-660.

Kim, W. and Hecht, M.H. (2005) Sequence determinants of enhanced amyloidogenicity of Alzheimer A{beta}42 peptide relative to A{beta}40. *J Biol Chem*, **280**, 35069-35076.

King, J., Haase-Pettingell, C., Robinson, A.S., Speed, M. and Mitraki, A. (1996) Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. *Faseb J*, **10**, 57-66.

King, J. and Laemmli, U.K. (1971) Polypeptides of the tail fibres of bacteriophage T4. *J Mol Biol*, **62**, 465-477.

Kishimoto, A., Hasegawa, K., Suzuki, H., Taguchi, H., Namba, K. and Yoshida, M. (2004) beta-Helix is a likely core structure of yeast prion Sup35 amyloid fibers. *Biochem Biophys Res Commun*, **315**, 739-745.

Knaus, K.J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W.K. and Yee, V.C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol*, **8**, 770-774.

Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. and Bukau, B. (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature*, **419**, 171-174.

Kreisberg, J.F., Betts, S.D., Haase-Pettingell, C. and King, J. (2002) The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. *Protein Sci*, **11**, 820-830.

Kreisberg, J.F., Betts, S.D. and King, J. (2000) Beta-helix core packing within the triple-stranded oligomerization domain of the P22 tailspike. *Protein Sci*, **9**, 2338-2343.

Krishnan, R. and Lindquist, S.L. (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature*, **435**, 765-772.

Kuhnel, K., Veltel, S., Schlichting, I. and Wittinghofer, A. (2006) Crystal structure of the human retinitis pigmentosa 2 protein and its interaction with Arl3. *Structure*, **14**, 367-378.

Kuroda, Y. and Kim, P.S. (2000) Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. *J Mol Biol*, **298**, 493-501.

Lander, G.C., Tang, L., Casjens, S.R., Gilcrease, E.B., Prevelige, P., Poliakov, A., Potter, C.S., Carragher, B. and Johnson, J.E. (2006) The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science*, **312**, 1791-1795.

Laurents, D.V. and Baldwin, R.L. (1998) Protein folding: matching theory and experiment. *Biophys J*, **75**, 428-434.

Lazo, N.D. and Downing, D.T. (1998) Amyloid fibrils may be assembled from beta-helical protofibrils. *Biochemistry*, **37**, 1731-1735.

Levinthal, C. (1968) Are there pathways for protein folding? *J Chem Phys*, **65**, 44-45.

Levinthal, C. (1969) How to fold graciously. University of Illinois Press, Urbana, pp. 22-24.

Liu, Y. and Eisenberg, D. (2002) 3D domain swapping: as domains continue to swap. *Protein Sci*, **11**, 1285-1299.

Liu, Y., Gotte, G., Libonati, M. and Eisenberg, D. (2001) A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol*, **8**, 211-214.

Liu, Y., Gotte, G., Libonati, M. and Eisenberg, D. (2002) Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci*, **11**, 371-380.

Liu, Y., Hart, P.J., Schlunegger, M.P. and Eisenberg, D. (1998) The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution. *Proc Natl Acad Sci U S A*, **95**, 3437-3442.

Lockless, S.W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295-299.

Lomas, D.A., Belorgey, D., Mallya, M., Miranda, E., Kinghorn, K.J., Sharp, L.K., Phillips, R.L., Page, R., Robertson, A.S. and Crowther, D.C. (2005) Molecular mousetraps and the serpinopathies. *Biochem Soc Trans*, **33**, 321-330.

Lombardi, A., Bryson, J.W. and DeGrado, W.F. (1996) De novo design of heterotrimeric coiled coils. *Biopolymers*, **40**, 495-504.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D. and Riek, R. (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A*, **102**, 17342-17347.

Maachupalli-Reddy, J., Kelley, B.D. and De Bernardez Clark, E. (1997) Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. *Biotechnol Prog*, **13**, 144-150.

Margittai, M. and Langen, R. (2006) Spin labeling analysis of amyloids and other protein aggregates. *Methods Enzymol*, **413**, 122-139.

McDonnell, A.V., Menke, M., Palmer, N., King, J., Cowen, L. and Berger, B. (2006) Fold recognition and accurate sequence-structure alignment of sequences directing beta-sheet proteins. *Proteins*, **63**, 976-985.

Michel, G., Helbert, W., Kahn, R., Dideberg, O. and Kloareg, B. (2003) The structural bases of the processive degradation of iota-carrageenan, a main cell wall polysaccharide of red algae. *J Mol Biol*, **334**, 421-433.

Miller, S., Schuler, B. and Seckler, R. (1998a) A reversibly unfolding fragment of P22 tailspike protein with native structure: the isolated beta-helix domain. *Biochemistry*, **37**, 9160-9168.

Miller, S., Schuler, B. and Seckler, R. (1998b) Phage P22 tailspike protein: removal of head-binding domain unmasks effects of folding mutations on native-state thermal stability. *Protein Sci*, **7**, 2223-2232.

Miller, S.R., Sekijima, Y. and Kelly, J.W. (2004) Native state stabilization by NSAIDs inhibits transthyretin amyloidogenesis from the most common familial disease variants. *Lab Invest*, **84**, 545-552.

Minton, A.P. (2006) How can biochemical reactions within cells differ from those in test tubes? *J Cell Sci*, **119**, 2863-2869.

Miranda, E. and Lomas, D.A. (2006) Neuroserpin: a serpin to think about. *Cell Mol Life Sci*, **63**, 709-722.

Mitraki, A., Danner, M., King, J. and Seckler, R. (1993) Temperature-sensitive mutations and second-site suppressor substitutions affect folding of the P22 tailspike protein in vitro. *J Biol Chem*, **268**, 20071-20075.

Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J. and King, J. (1991) Global suppression of protein folding defects and inclusion body formation. *Science*, **253**, 54-58.

Mukhopadhyay, S., Krishnan, R., Lemke, E.A., Lindquist, S. and Deniz, A.A. (2007) A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A*, **104**, 2649-2654.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R. and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**, 773-778.

Nguyen, J., Baldwin, M.A., Cohen, F.E. and Prusiner, S.B. (1995) Prion protein peptides induce alpha-helix to beta-sheet conformational transitions. *Biochemistry*, **34**, 4186-4192.

Nilsson, M.R. (2004) Techniques to study amyloid fibril formation in vitro. *Methods*, **34**, 151-160.

Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M. and Goldman, A. (2004) The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *Embo J*, **23**, 701-711.

O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, **254**, 539-544.

O'Shea, E.K., Lumb, K.J. and Kim, P.S. (1993) Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr Biol*, **3**, 658-667.

Otto, B.R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J.G., Mizutani, K., Park, S.Y. and Tame, J.R. (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli. *J Biol Chem*, **280**, 17339-17345.

Parent, K.N., Ranaghan, M.J. and Teschke, C.M. (2004) A second-site suppressor of a folding defect functions via interactions with a chaperone network to improve folding and assembly in vivo. *Mol Microbiol*, **54**, 1036-1050.

Perutz, M.F., Finch, J.T., Berriman, J. and Lesk, A. (2002) Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci U S A*, **99**, 5591-5595.

Petersen, T.N., Kauppinen, S. and Larsen, S. (1997) The crystal structure of rhamnogalacturonase A from Aspergillus aculeatus: a right-handed parallel beta helix. *Structure*, **5**, 533-544.

Petty, S.A., Adalsteinsson, T. and Decatur, S.M. (2005) Correlations among morphology, beta-sheet stability, and molecular structure in prion peptide aggregates. *Biochemistry*, **44**, 4720-4726.

Petty, S.A. and Decatur, S.M. (2005a) Experimental evidence for the reorganization of beta-strands within aggregates of the Abeta(16-22) peptide. *J Am Chem Soc*, **127**, 13488-13489.

Petty, S.A. and Decatur, S.M. (2005b) Intersheet rearrangement of polypeptides during nucleation of {beta}-sheet aggregates. *Proc Natl Acad Sci U S A*, **102**, 14272-14277.

Pickersgill, R., Harris, G., Lo Leggio, L., Mayans, O. and Jenkins, J. (1998) Superfamilies: the 4/7 superfamily of beta alpha-barrel glycosidases and the right-handed parallel beta-helix superfamily. *Biochem Soc Trans*, **26**, 190-198.

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

Pope, W.H., Haase-Pettingell, C. and King, J. (2004) Protein folding failure sets high-temperature limit on growth of phage P22 in Salmonella enterica serovar Typhimurium. *Appl Environ Microbiol*, **70**, 4840-4847.

Prusiner, S.B. (1998) Prions. *Proc Natl Acad Sci U S A*, **95**, 13363-13383.

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E. and Razgulyaev, O.I. (1990) Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett*, **262**, 20-24.

Qu, B.H., Strickland, E. and Thomas, P.J. (1997) Cystic fibrosis: a disease of altered protein folding. *J Bioenerg Biomembr*, **29**, 483-490.

Raetz, C.R. and Roderick, S.L. (1995) A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science*, **270**, 997-1000.

Richardson, J.S. and Richardson, D.C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A*, **99**, 2754-2759.

Rinas, U., Hoffmann, F., Betiku, E., Estape, D. and Marten, S. (2007) Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli. *J Biotechnol*, **127**, 244-257.

Robinson, A.S. and King, J. (1997) Disulphide-bonded intermediate on the folding and assembly pathway of a non-disulphide bonded protein. *Nat Struct Biol*, **4**, 450-455.

Roggenkamp, A., Ackermann, N., Jacobi, C.A., Truelzsch, K., Hoffmann, H. and Heesemann, J. (2003) Molecular analysis of transport and oligomerization of the Yersinia enterocolitica adhesin YadA. *J Bacteriol*, **185**, 3735-3744.

Rotondi, K.S. and Gierasch, L.M. (2003a) Local sequence information in cellular retinoic acid-binding protein I: specific residue roles in beta-turns. *Biopolymers*, **71**, 638-651.

Rotondi, K.S. and Gierasch, L.M. (2003b) Role of local sequence in the folding of cellular retinoic abinding protein I: structural propensities of reverse turns. *Biochemistry*, **42**, 7976-7985.

Rotondi, K.S., Rotondi, L.F. and Gierasch, L.M. (2003) Native structural propensity in cellular retinoic acid-binding protein I 64-88: the role of locally encoded structure in the folding of a beta-barrel protein. *Biophys Chem*, **100**, 421-436.

Rousseau, F., Schymkowitz, J.W. and Itzhaki, L.S. (2003) The unfolding story of three-dimensional domain swapping. *Structure*, **11**, 243-251.

Rousseau, F., Schymkowitz, J.W., Wilkinson, H.R. and Itzhaki, L.S. (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci U S A*, **98**, 5596-5601.

Rowe, S.M., Miller, S. and Sorscher, E.J. (2005) Cystic fibrosis. *N Engl J Med*, **352**, 1992-2001.

Saint-Jean, A.P., Phillips, K.R., Creighton, D.J. and Stone, M.J. (1998) Active monomeric and dimeric forms of Pseudomonas putida glyoxalase I: evidence for 3D domain swapping. *Biochemistry*, **37**, 10345-10353.

Sambashivan, S., Liu, Y., Sawaya, M.R., Gingery, M. and Eisenberg, D. (2005) Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature*, **437**, 266-269.

Sanders, A., Jeremy Craven, C., Higgins, L.D., Giannini, S., Conroy, M.J., Hounslow, A.M., Waltho, J.P. and Staniforth, R.A. (2004) Cystatin forms a tetramer through structural rearrangement of domain-swapped dimers prior to amyloidogenesis. *J Mol Biol*, **336**, 165-178.

Sauer, R.T. (1996) Protein folding from a combinatorial perspective. *Fold Des*, **1**, R27-30.

Sauer, R.T., Krovatin, W., Poteete, A.R. and Berget, P.B. (1982) Phage P22 tail protein: gene and amino acid sequence. *Biochemistry*, **21**, 5811-5815.

Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., Madsen, A.O., Riekel, C. and Eisenberg, D. (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453-457.

Scavetta, R.D., Herron, S.R., Hotchkiss, A.T., Kita, N., Keen, N.T., Benen, J.A., Kester, H.C., Visser, J. and Jurnak, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell*, **11**, 1081-1092.

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I. and Ban, N. (2006) Structure of the E. coli signal recognition particle bound to a translating ribosome. *Nature*, **444**, 503-506.

Scheibel, T., Bloom, J. and Lindquist, S.L. (2004) The elongation of yeast prion fibers involves separable steps of association and conversion. *Proc Natl Acad Sci U S A*, **101**, 2287-2292.

Schuler, B., Rachel, R. and Seckler, R. (1999) Formation of fibrous aggregates from a non-native intermediate: the isolated P22 tailspike beta-helix domain. *J Biol Chem*, **274**, 18589-18596.

Schuler, B. and Seckler, R. (1998) P22 tailspike folding mutants revisited: effects on the thermodynamic stability of the isolated beta-helix domain. *J Mol Biol*, **281**, 227-234.

Schwahn, U., Lenzner, S., Dong, J., Feil, S., Hinzmann, B., van Duijnhoven, G., Kirschner, R., Hemberger, M., Bergen, A.A., Rosenberg, T., Pinckers, A.J., Fundele, R., Rosenthal, A., Cremers, F.P., Ropers, H.H. and Berger, W. (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet*, **19**, 327-332.

Schweers, O., Schonbrunn-Hanebeck, E., Marx, A. and Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J Biol Chem*, **269**, 24290-24297.

Seckler, R. (1998) Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J Struct Biol*, **122**, 216-222.

Seckler, R., Fuchs, A., King, J. and Jaenicke, R. (1989) Reconstitution of the thermostable trimeric phage P22 tailspike protein from denatured chains in vitro. *J Biol Chem*, **264**, 11750-11753.

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F. and Lindquist, S.L. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, **289**, 1317-1321.

Silva, R.A., Barber-Armstrong, W. and Decatur, S.M. (2003) The organization and assembly of a beta-sheet formed by a prion peptide in solution: an isotope-edited FTIR study. *J Am Chem Soc*, **125**, 13674-13675.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J. and Whisstock, J.C. (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem*, **276**, 33293-33296.

Simkovsky, R. and King, J. (2006) An elongated spine of buried core residues necessary for in vivo folding of the parallel beta-helix of P22 tailspike adhesin. *Proc Natl Acad Sci U S A*, **103**, 3575-3580.

Smith, D.H. and King, J. (1981) Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail spike protein. III. Intensive polypeptide chains synthesized at 39 degrees C. *J Mol Biol*, **145**, 653-676.

Speed, M.A., Morshead, T., Wang, D.I. and King, J. (1997) Conformation of P22 tailspike folding and aggregation intermediates probed by monoclonal antibodies. *Protein Sci*, **6**, 99-108.

Speed, M.A., Wang, D.I. and King, J. (1995) Multimeric intermediates in the pathway to the aggregated inclusion body state for P22 tailspike polypeptide chains. *Protein Sci*, **4**, 900-908.

Steinbacher, S., Baxa, U., Miller, S., Weintraub, A., Seckler, R. and Huber, R. (1996) Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. *Proc Natl Acad Sci U S A*, **93**, 10584-10588.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R. and Huber, R. (1997a) Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 A resolution, and the molecular basis of O-antigen recognition and cleavage. *J Mol Biol*, **267**, 865-880.

Steinbacher, S., Miller, S., Baxa, U., Weintraub, A. and Seckler, R. (1997b) Interaction of Salmonella phage P22 with its O-antigen receptor studied by X-ray crystallography. *Biol Chem*, **378**, 337-343.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R. and Reinemer, P. (1994) Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science*, **265**, 383-386.

Sturtevant, J.M., Yu, M.H., Haase-Pettingell, C. and King, J. (1989) Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. *J Biol Chem*, **264**, 10693-10698.

Sunde, M. and Blake, C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem*, **50**, 123-159.

Tessier, P.M. and Lindquist, S. (2007) Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature*, **447**, 556-561.

Thanassi, D.G., Stathopoulos, C., Karkal, A. and Li, H. (2005) Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of gram-negative bacteria (review). *Mol Membr Biol*, **22**, 63-72.

Thompson, J.R., Bratt, J.M. and Banaszak, L.J. (1995) Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular beta-sheet. *J Mol Biol*, **252**, 433-446.

True, H.L., Berlin, I. and Lindquist, S.L. (2004) Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature*, **431**, 184-187.

Valax, P. and Georgiou, G. (1993) Molecular characterization of beta-lactamase inclusion bodies produced in Escherichia coli. 1. Composition. *Biotechnol Prog*, **9**, 539-547.

Varley, P., Gronenborn, A.M., Christensen, H., Wingfield, P.T., Pain, R.H. and Clore, G.M. (1993) Kinetics of folding of the all-beta sheet protein interleukin-1 beta. *Science*, **260**, 1110-1113.

Viguera, A.R., Serrano, L. and Wilmanns, M. (1996) Different folding transition states may result in the same native structure. *Nat Struct Biol*, **3**, 874-880.

Vincent, J.J., Tai, C.H., Sathyanarayana, B.K. and Lee, B. (2005) Assessment of CASP6 predictions for new and nearly new fold targets. *Proteins*, **61 Suppl 7**, 67-83.

Weigele, P.R., Haase-Pettingell, C., Campbell, P.G., Gossard, D.C. and King, J. (2005) Stalled folding mutants in the triple beta-helix domain of the phage P22 tailspike adhesin. *J Mol Biol*, **354**, 1103-1117.

Weigele, P.R., Scanlon, E. and King, J. (2003) Homotrimeric, beta-stranded viral adhesins and tail proteins. *J Bacteriol*, **185**, 4022-4030.

Wetzel, R. (2002) Ideas of order for amyloid fibril structure. *Structure (Camb)*, **10**, 1031-1036.

Whitesell, L. and Lindquist, S.L. (2005) HSP90 and the chaperoning of cancer. *Nat Rev Cancer*, **5**, 761-772.

Wilczynska, M., Fa, M., Ohlsson, P.I. and Ny, T. (1995) The inhibition mechanism of serpins. Evidence that the mobile reactive center loop is cleaved in the native protease-inhibitor complex. *J Biol Chem*, **270**, 29652-29655.

Williams, A.D., Portelius, E., Kheterpal, I., Guo, J.T., Cook, K.D., Xu, Y. and Wetzel, R. (2004) Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J Mol Biol*, **335**, 833-842.

Williams, A.H., Immormino, R.M., Gewirth, D.T. and Raetz, C.R. (2006) Structure of UDP-N-acetylglucosamine acyltransferase with a bound antibacterial pentadecapeptide. *Proc Natl Acad Sci U S A*, **103**, 10877-10882.

Wishner, B.C., Ward, K.B., Lattman, E.E. and Love, W.E. (1975) Crystal structure of sickle-cell deoxyhemoglobin at 5 A resolution. *J Mol Biol*, **98**, 179-194.

Woolfson, D.N. and Alber, T. (1995) Predicting oligomerization states of coiled coils. *Protein Sci*, **4**, 1596-1607.

Woolhead, C.A., Johnson, A.E. and Bernstein, H.D. (2006) Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell*, **22**, 587-598.

Wu, Y., Swulius, M.T., Moremen, K.W. and Sifers, R.N. (2003) Elucidation of the molecular logic by which misfolded alpha 1-antitrypsin is preferentially selected for degradation. *Proc Natl Acad Sci U S A*, **100**, 8229-8234.

Yoder, M.D., Keen, N.T. and Jurnak, F. (1993a) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science*, **260**, 1503-1507.

Yoder, M.D., Lietzke, S.E. and Jurnak, F. (1993b) Unusual structural features in the parallel beta-helix in pectate lyases. *Structure*, **1**, 241-251.

Yu, M.H. and King, J. (1984) Single amino acid substitutions influencing the folding pathway of the phage P22 tail spike endorhamnosidase. *Proc Natl Acad Sci U S A*, **81**, 6584-6588.

Yu, M.H., Weissman, J.S. and Kim, P.S. (1995) Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J Mol Biol*, **249**, 388-397.

Zhang, H., Kaneko, K., Nguyen, J.T., Livshits, T.L., Baldwin, M.A., Cohen, F.E., James, T.L. and Prusiner, S.B. (1995) Conformational transitions in peptides containing two putative alpha-helices of the prion protein. *J Mol Biol*, **250**, 514-526.

Zhou, A., Stein, P.E., Huntington, J.A., Sivasothy, P., Lomas, D.A. and Carrell, R.W. (2004) How small peptides block and reverse serpin polymerisation. *J Mol Biol*, **342**, 931-941.

CHAPTER TEN:

APPENDICES

APPENDIX A. GELANALYZER PLUG-IN FOR IMAGEJ

Commercial 1D gel analysis software is readily available from a number of commercial suppliers. These software packages, such as ImageQuant (MD), allow for the quantification of optical densities of bands in electrophoresis gels, an activity that proved indispensable in this thesis. However, many of these programs suffer from a number of limitations that restrict their usefulness, such as operating system dependencies, the expense of the software, the inability to correct mistakes, and limits on the number of samples that can be analyzed at one time. Many of these software programs are laborious and time consuming, particularly when attempting to analyze a 102-lane gel image.

In order to overcome many of the drawbacks inherent in commercially available gel analysis programs, I developed the GelAnalyzer plug-in for the freely available image processing software ImageJ (Wayne Rasband, National Institutes of Health, USA). Although ImageJ has its own methods for performing optical densitometry analysis, these routines suffer from many of the same problems as the commercial software. For example, there is currently no way to undo mistakes without having to start the process over entirely. ImageJ also only allows for the selection of perfectly rectangular lanes, and does not allow for lanes that shift, which is a common occurrence in scanned gel images. GelAnalyzer was specifically developed to take advantage of the established framework of the Java-based ImageJ program while making the process of gel analysis straight forward, flexible, and accurate.

Like ImageJ, the plug-in was written in the Java programming language, making it cross-platform. The plug-in is also freely distributed and can be installed in one step, making it readily accessible to any individual interested in performing gel analysis. The source code for GelAnalyzer is packaged with the program, allowing software developers the option of expanding GelAnalyzer or using its methods as a library for other programs.

Once a gel image is opened in the ImageJ program, the plug-in can be started from ImageJ's plug-in menu. The plug-in displays a graphical user interface (GUI) for the user, which takes the user through the five basic steps of analyzing a gel: orientation of the image, calibration of the image, marking lanes, subtracting background, and identifying and analyzing band peaks. At any point in the process, the user can save whatever progress has been performed on the current image or load progress from a

previous analysis run. All selected values for each step of the analysis are saved in the text-formatted progress file.

1. Image Orientation

Although the GUI is the main control center for the analysis process, the user can still interact with the image using many of the ImageJ tools. This is readily observable in the first phase of gel analysis, image orientation. The image orientation panel provides two main methods of rotating the image, either by automated detection of features or by manual inputting rotation values.

Two options are currently implemented for automated orientation of the gel image. The first is based on the detection of a step tablet, while the second attempts to recognize the gel itself. Neither of these methods are guaranteed to provide the desired results, but the reliability of their detection algorithms may increase with time. Currently, the step tablet recognition algorithm, which relies on identifying the darkest and lightest spots on the image and predicting where the corners of the tablet are based on these values, appears to perform more reliably than the gel detection algorithm. Gel detection implements a modified Sobel filter to detect band edges. The algorithm then attempts to calculate the median directionality of detected bands, and rotate the gel accordingly.

Manual orientation has so far proven the most reliable method for rotating for gel analysis. The user can input a desired angle of rotation and tell the program if the image should be flipped in the vertical or the horizontal directions. The user may use ImageJ's line selection tool to draw a line over a desired image feature and ask the plug-in to make the line selection horizontal. The program will automatically calculate the desired angle of rotation with respect to the original image's orientation and report that angle back to the user.

2. Image Calibration

Image calibration is a necessary step for reliable optical densitometry evaluations. Images may be scanned using different scanners or different scanner settings, which changes how a black or white pixel value in an image is related to absolute values of optical density. This is also important because the relationship between optical density and pixel value is rarely a linear function, but more often is a logarithmic function.

If an image is not scanned with a step tablet or without a step tablet of known optical density values, then the user has the option of calibrating the image using ImageJ's built-in calibration function "Uncalibrated OD". Although this is not the most accurate method of calibration, it does at least attempt an exponential relationship between optical density and pixel value.

Alternatively, the user may interact with the step tablet outlining tool that is displayed on the image. This tool is a modified and enhanced version of ImageJ's region of interest (Roi) selection boxes. It can therefore be resized and relocated at will. The height and the amount of sampling to calculate pixel values can be adjusted in the GUI's calibration panel. The number of steps as well as each step's width and OD value can be adjusted using the displayed table interface. The calculated average pixel value for each step is automatically recalculated any time the step tablet is changed. Finally, the type of function that will be used to map pixel values to optical densitometry values can be selected from amongst 12 built-in options. However, every time the step tablet is adjusted and the average pixel values of steps are calculated, the fit function yielding the best fit is automatically calculated and displayed in its selector.

Once the user has outlined the step tablet, selected the appropriate fit function, and pressed the calibrate button, the image's pixel values are calibrated according to the inner mechanisms of ImageJ's calibration and function. Pixel values can be automatically transformed into optical density values.

3. Marking Lanes

Once the image is calibrated, the gel lanes are ready to be marked. Although the gel selection panel does not currently have methods for automatic lane or band detection, the java programming interface allows for such methods to be easily incorporated into the program if they are developed.

Currently, gel lanes must be traced manually. A set of templates for common gel sizes and numbers of lanes are incorporated into the program for easy selection. Like the step tablet selection tool, the gel selection tool is an extended version of ImageJ's roi selectors. Therefore, the user can resize or translate the gel in the image window. Alternatively, the lane width, lane length, and directionality of the gel can be determined in GelAnalyzer's GUI. It is an important component of gel analysis that all lanes have the same width because the average optical density is calculated of a band is calculated as the average OD across the width of the lane. In order for all lanes to share the same amount of noise in this calculation, lane widths are kept constant across all lanes.

Lanes can be individually placed using the GUI's input table. Lanes can also be named, either by manually typing in the lane names or by loading the lane names from a comma-delimited file. In order to correct for possible lane shifting, lanes can be manipulated individually on the gel image once the "Individual Lanes" adjustment mode is selected. The lane selection tool is a spline tool, where the center of the lane is determined by a spline through a set of node points. This allows for curvature in the lane and let's the user define specific locations that the center of the lane must pass through. Nodes can be added or removed by holding down the "alt" key while clicking on the image window. The lane selection tool is also a type of roi selector, and thus it can be translated at will.

It is important to note that although the spline allows for curvature in a lane, the profile plot of the lane is calculated as the average OD over a given width of pixels at a specific pixel height in the image. The profile plot does not adjust its calculation to take into account the angle of curvature in the lane.

4. Background Subtraction

Once lanes have been selected, the GelAnalyzer plug-in automatically calculates the plot profile of the lane as the average optical density at each integral pixel position along the height of the lane. The user is then offered the option to correct individual lanes by a specified background lane. This is a useful tool if the user desires to remove the background staining of the gel itself, based on a lane not loaded with any sample, or if the user wishes to subtract the background bands of a control sample.

Each lane or all lanes can be selected as background, as being corrected by the background lane, or for not being corrected at all. In the current version of the program, if multiple lanes are chosen as background, then the average profile plot of these lanes is calculated as the background to use for subtractions. The background selection tool allows for the input of a horizontal offset of the current lane in comparison to the background profile plot. This enables the user to correct for smiles in the gel that causes the same bands in different lanes to run at slightly different positions from each other. Although this kind of correction would ideally allow for horizontal scaling to align peaks with each other, the program does not currently implement that ability. Offsets can be calculated by an algorithm that attempts to match up peaks, though this automated algorithm is not guaranteed to work often offsets lanes about one to two pixels off from the ideal horizontal offset.

Since some lanes may stain differently than other lanes, scalars can be applied to each lane in order to have the baselines of the background match the baselines of the sample lane. An algorithm is provided that calculates the ideal scalar for matching the median OD values of the two lanes. This algorithm was used so as to attempt to match the baselines of the lanes, which ideally are the most prevalent OD value in each lane.

When determining the background, scalars, and offsets of each lane, profile plots are shown depicting the relative values of the raw profile plot, the background profile plot, and the final adjusted profile plot if the lane is selected for background correction. These plots are updated any time a change is made to the lane's offset or scalar in order to visually aid the user.

5. Peak Analysis

Having defined lanes and defined whether sample lanes are to be corrected by background lanes, the user can analyze the individual peaks in each lane. When this panel of the GelAnalyzer GUI is first opened or anytime the lane and background definitions are changed, an automated peak detection algorithm run in an attempt to detect all of the peaks. This algorithm scans the profile plot for any local maxima in the plot and defines those positions as peaks. This is not ideal, in that smoothing of the profile plot may produce fewer spurious peaks

For each lane, the user can specify a variable number of peaks and can name each peak individually. Each peak can be selected graphically or a horizontal apex position can be typed into the peaks table. Once a new apex value is given to the program, it will automatically recalculate where the nearest apex is for the selected peak and will try to detect the upper and lower limits of the peak. The user can also manually adjust the bounds of the peak. Peaks bounds are displayed in the gel image so that the user can compare peak locations across lanes and can visually see where in the gel the peak is located. The number of peaks in a lane, the peaks' names, the peaks' values, and the peaks' bounds can be individually selected for each lane or can be applied to all lanes.

A baseline value can also be adjusted for calculating the integrated optical density of the peak. The baseline can be assigned a specific numerical value to set as a flat baseline for the entire profile peak. Most often, this baseline would be set to 0.0. Alternatively, the user can set the baseline mode to be sloped, which will calculate a linear sloped baseline from the minimum bound of the peak to the maximum bound of the peak. This baseline can be adjusted individually for each lane, or can be set for all lanes in the gel.

Once the peaks are defined, the positive integrated optical density between the profile plot and the baseline is automatically calculated. This represents the optical density for that protein band, while the apex value represents the distance migrated. These values can be saved in a comma-delimited file that can be read by most data analysis or spreadsheet software.

APPENDIX B. HIS-TAGGED TAILSPIKE CONSTRUCTS

When β-helix mutants were initially investigated for their ability to fold *in vitro* in the context of the full length tailspike protein, his-tagged constructs of these mutants were prepared. Dr. Peter Weigele and Phil Campbell cloned the wild-type *gene 9* sequence into a pET21b vector using the BamHI and NdeI restriction digest cloning sites. The use of these cloning sites added a total of 22 amino acids to the end of the protein. DNA sequencing confirmed the addition of the amino acid sequence "ADPNSSSVDKLAAALEHHHHHH" to the C-terminus of the protein. This plasmid, termed pET(TSPK:H6), was subsequently used as a template for site directed mutagenesis, as performed in Chapter 2. Resulting plasmids were confirmed by sequencing and were transformed into BL21(DE3) cells in preparation for over-expression.

Over-expressed proteins were purified under denaturing conditions using a Ni-NTA metal affinity column to purify the his-tagged tailspike proteins. Proteins were purified to >90% purity. As a control for the purification of proteins under denaturing conditions, wild-type his-tagged tailspike proteins were purified under native conditions from the supernatant fraction of the same over-expression culture used to produce the wild-type proteins for denaturing purification. Natively purified wild-type his-tagged chains were named WN, while wild-type his-tagged chains purified under denaturing conditions were called WD.

After an one hour incubation in 8 M urea at pH 3, proteins were examined for their ability to refold as performed in Chapter 5. While WN was able to refold, none of the other tailspike sequences were able to refold under the tested conditions. In this experiment, WD was also unable to fold to a native trimeric form and instead aggregated as determined by native gel electrophoresis.

Various experimental conditions were adjusted in an attempt to get the WD sample to refold like the WN sample. As a control for this experiment, wild-type tailspike protein that lacked the his-tag was also simultaneously refolded. Various purification and unfolding buffers were examined for their effects on the tailspike chains. In the initial

purification, 7 M urea of varying pH from 8 to 4.6 were used to bind, wash, and elute the tailspike protein from the nickel column. The denaturing purification buffers were adjusted to elute the proteins with varying concentrations of immidazole at pH 8. The denaturing buffers were further adjusted to replace urea with 6 M guanidine hydrochloride. Independent of the buffer conditions used to purify the wild-type his-tagged tailspike protein under denaturing conditions the WD sample would not refold.

The unfolding buffer used before dilution with refolding buffer was also adjusted. Initial investigations used an 8 M urea solution at pH 3.0, as the natively folded tailspike protein will not unfold at neutral pH even in high concentrations of urea. Further investigations used the neutral pH buffer used in Chapter 5. Independent of the unfolding buffer used, the WD sample did not refold. This is in contrast to the data presented in Chapter 5, which demonstrated that the wild-type full length tailspike protein was able to refold out of aggregated inclusion bodies.

Further experimentation demonstrated that the inability to refold was not rescued by the presence of DTT or BME. The introduction of 1,10-phenanthroline, an iron chelator, or the additon of iron sulfate did not influence the folding of the WD or WN samples. Experimentation did not reveal any conditions under which the WD protein samples were able to refold to a trimeric tailspike protein.

In contrast, it was discovered that incubation in 6 M urea at pH 3.0 at 10ºC for an extended period of time, approximately 5 days, caused the WN sample to no longer be able to refold. A control reaction of natively purified, wild-type tailspike lacking the histidyl tag was simultaneously examined. After 5 days of incubation in an acidic urea solution, the wild-type tailspike lacking a his-tag was still able to refold as it had after one hour of incubation. The results indicate that the C-terminal 22 amino acid tag that includes six histidines was sensitive to the presence of denaturants in a time-dependent manner.

We hypothesized that the observed sensitivity was likely due to chemical modifications of the protein, most likely in the C-terminal tag. The MIT biopolymers group analyzed WN and WD samples by mass spectroscopy to determine if any

molecular weight differences existed between proteins in the two stock samples. Initial Maldi analysis revealed that a molecular weight difference of 154 daltons. Trypsin digest followed by ion spray or Maldi demonstrated that two extra bands resulted from digestion of the WN sample that were not present in the WD samples. The molecular weights of these bands, however, did not map to expected masses from *in silico* digests.

Upon attempting to purify these tailspike proteins by HPLC in preparation for repeating the mass spectrometry experiments, samples behaved unexpectedly. Samples either stuck to the column or samples eluted in broad and mixed peaks.

When the above mass spectrometry experiments were repeated with stock sample that had been refrigerating for nearly two months, no mass difference could be detected.

To date, it is unclear what modifications or interactions may have been occurring to cause the observed inability to refold. The only clear result is that the 22 amino acid C-terminal tag was likely responsible in a manner dependent upon long time exposure to denaturant.