Title
Identifying the Structural Determinants of Extreme Folding and Unfolding Barriers

Author
Kelch, Brian A

Publication Date
2007-04-02

Peer reviewed|Thesis/dissertation

Identifying the Structural Determinants of Extreme Folding and Unfolding Barriers

by

Brian A. Kelch

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

I dedicate this thesis to

My father, Anthony W. Kelch,

A man who asks for nothing, but deserves everything.

Acknowledgements

I would like to thank David Agard for his guidance and advice throughout all my years at UCSF. His input has been instrumental in my research as well as my development as a scientist. I also wanted to thank Dave teaching me how to communicate science in a clear and interesting way. He put in a tremendous effort both into the oral and written forms of communication and, while I still have much to learn, I feel that this has been one of the areas in which I have progressed the most in graduate school. I had a great time working for Dave and will miss his input, sharp intellect and energetic enthusiasm.

My thesis committee was also extremely helpful in getting me through grad school. Jonathan Weissman is great scientist who thinks deeply about how folding affects biological processes, which has directed a lot of my thinking from pure biophysical concepts towards how the physics determines biological functionality. Susan Marqusee has been a fantastic mentor whose expertise in protein folding was instrumental in the direction of my progress throughout grad school. I'd also like to thank her for her mentoring skills, guiding my professional as well as scientific development.

The Agard Lab was a fantastic place to perform science for the 7.5 years I was there. There are too many people in the Agard lab to list them all and their assistance in shaping my scientific career; if you are reading this, which most of you will never do, I'd like to express a hearty "Thank you." I have been honored to be able to work with such a warm, friendly and helpful group of people.

That being said, I'd like to thank these specific lab members, both past and present, for their extra-special help thoughout my graduate career.

First, I'd like to thank Stephanie Truhlar. Steph and I came into the Agard lab together as rotons together and were, right away, close friends. Steph managed to put up with my hijinks (even while sharing a desk for several months!) Throughout grad school, we had tough times, but the friendship between Steph and I was stronger than anything grad school threw at us. I thank her deeply for her understanding, friendship, and sound scientific advice.

Luke Rice is a post-doc who has been instrumental in my development as a scientist. Luke is amazingly intelligent and is willing to expend his time and energy thinking about the scientific problems of others in the lab. It's almost as if Luke is an extra PI who is actually in the lab. His help spans teaching me how to deal with technical crystallographic issues to thorny interpretation of results. I thank him enough for the help he has provided.

Cynthia Fuhrmann, who, as a lab member (see below for her role as a wife), was extremely helpful in assisting me with crystallographic data collection and processing. She dealt with my frenetic late-night data collections at the Stinkytron with a Buddha-like calm, dishing out sound advice all the while. Her identification of the distorted phenylalanine was instrumental in guiding that aspect of my graduate career, and her assistance with writing papers was extremely helpful in the hard, long process of improving my written communication skills. Especially in rooting out my tendency towards redundancy, wordiness and repeating myself.

I also would like to thank the many friends that I have made in graduate school, for grounding me during this long, arduous struggle. Adam Douglass was my roommate for 3 or 4 years and we got along fantastically well, despite the fact that he actually had to live with me. We had a blast living together, creating greasy, bacony messes (the bacon candle, fry parties, etc.) and drinking till the wee hours. I'm really grateful to Adam for the friendship, laughs and good times.

Janet Iwasa was almost as much a part of our household as either Adam or I. Janet's an amazing person with a treasure trove of hiddent talents. Who, other than Janet, raises their own Angora rabbit so that they can harvest the fur, spin it into yarn, and knit hats and scarves from it? Who, other than Janet, can learn how to make a full-on giant wedding cake with all the sugar flower trimmings within a few months? I'm always in awe of Janet's talents. Her friendship has been indispensable as well. I wish her and Adam all the best in their new home in Boston.

Meter and his wife Sonia were also incredible friends throughout the grad school years. They are two of the most caring people that I have met and were really great sources of support. I always loved hanging out with both of them, whether it was camping, going to a concert or just hanging out. It was always a good time.

Ethan and Quincey were fantastic friends as well. Cynthia and I always looked forward to our Friday evening dinner with them. We always had a blast and I'm looking forward to more good times (since they're the only ones who haven't left SF!)

I'd also like to thank all my other friends Jen Berman, L-train, Scott, Laura, Brian Margolin, Alex Kelly, (taste the) Nilesh, and all the others who I have forgotten here.

Their friendship and laughter made grad school bearable. Without them, I don't think I would have made it through this meat-grinder called grad school. Thanks.

My family is an important part of my life and I want to thank all of them for the love and support through the years. There's too many family members to mention (even immediate family!), so I'll just say "Thank you!"

I have to give extra-special thanks to my parents, Tony and Maureen Kelch (known to me as "Dad" and "Mom".) Their love and support have been instrumental in my development, all the way from infant to now as a grown man. I'd like to thank my father for the support of my love of science from an early age. I remember being in second grade and Dad explaining to me, in detail, Newton's laws of motion. I credit my father for getting instilling in me the desire to 'become a scientist when I grow up.' My mother has been very caring, loving and supportive all throughout my life and especially these last few years. It was one of the happiest moments to finally have an answer to her question "When are you going to be done with grad school?" I credit my mother for instilling in me the desire to be a merry prankster, which I'm sure the rest of the Agard lab is upset about.

Finally I'd like to thank my wife Cynthia. The best aspect of grad school here at UCSF has been my time with her. She is the love of my life and my anchor when things go wrong (which is a constant in grad school.) Cynthia made me the happiest (and many would say luckiest) man alive when she married me in May 2005. I am eternally grateful for her love and support and am looking forward to spending our lives together as a family.

Abstract

Detailed knowledge of folding intermediate and transition state (TS) structures is critical for understanding protein folding mechanisms. For kinetically-stable proteins such as α-lytic protease (αLP) and its family members, their large free energy barrier to unfolding is central to their biological function. Thus their TS structure plays a crucial role in protein function. However, structural information regarding this important state has been completely lacking, mainly because standard techniques to probe TS structure are not realistically applicable for αLP. Therefore, I used the information embedded in the sequence of homologous proteases to discern the physical mechanisms by which kinetic stability can be modulated. This required experimental validation using various biophysical and biochemical techniques, such as mutagenesis, x-ray crystallography, and detailed kinetic analyses.

From these studies, I have shown that the conserved distortion of a sidechain significantly contributes to the destabilization of the TS for a large sub-class of αLP homologs. The strain from this deformation actually provides a biological advantage in that lifetime is greatly extended. This study was the first that shows that sidechain distortion has been shown to be used for a functional purpose and uncovers an unanticipated challenge for structural biology to identify potentially relevant distortions from high resolution structural studies.

My structural and kinetic analysis of a acid resistant αLP homolog, *Nocardiopsis alba* Protease (NAPase), identified the physical basis for this proteins acid stability, thus providing crucial structural information about unfolding mechanisms and leading to a model for the TS structure for these proteases. This study provided insight into the

evolutionary benefits of kinetic stability as a paradigm for generation of extremophilic behavior.

From a similar study of a thermophilic αLP homolog, *Thermobifida fusca* Protease A (TFPA), I identified a substructure of these proteases, termed the domain bridge, which is used to modulate the degree of kinetic stability. This study refined our model for the unfolding TS, in which the domain bridge undocks and unfolds allowing the two domains of the protease to separate, with the newly formed crevice filling with solvent. These studies represent the first physical understanding of the structural basis for kinetic stability.

List of Tables

List of Figures

Table of Contents

Chapter 1: Introduction

αLP is Kinetically Stable

It is axiomatic that most proteins spontaneously fold to their global free energy minimum and that this state corresponds to their native conformation(Anfinsen 1973), a growing number of proteins do not conform to this hypothesis. Many proteins, including nearly all extracellular bacterial proteases, require a covalently attached Pro domain for proper folding and maturation(Baker, Shiau et al. 1993; Bryan 2002). Alpha-Lytic Protease (αLP; Figure 1.1), a 198-residue serine protease of the trypsin superfamily secreted by the soil bacterium *Lysobacter enzymogenes*, is perhaps the best studied of these proteins(Fuhrmann 2003). αLP is synthesized with a 166-residue N-terminal Pro Region which has been shown to be necessary, either in *cis* as a continous polypeptide or in *trans* as a separate protein, for proper folding and secretion of the protease(Silen and Agard 1989; Silen, Frank et al. 1989).

To understand this *in vivo* folding requirement for the Pro Region, the folding landscape for αLP was determined *in vitro*(Baker, Sohl et al. 1992; Sohl, Jaswal et al. 1998)(Figure 1.2). When denatured αLP is diluted or dialyzed from denaturant, a stable molten globule intermediate forms(Baker, Sohl et al. 1992), which interconverts extremely slowly ($t_{1/2}$ ~2000 years) to the native state(Sohl, Jaswal et al. 1998). The rate of αLP unfolding, although extremely slow itself ($t_{1/2}$ ~1.2 years), is significantly faster than the folding rate, indicating that the αLP native state is less stable than Int by a remarkable 4 kcal/mol(Sohl, Jaswal et al. 1998). Therefore, the large unfolding free energy barrier is the sole mechanism of stability for αLP.

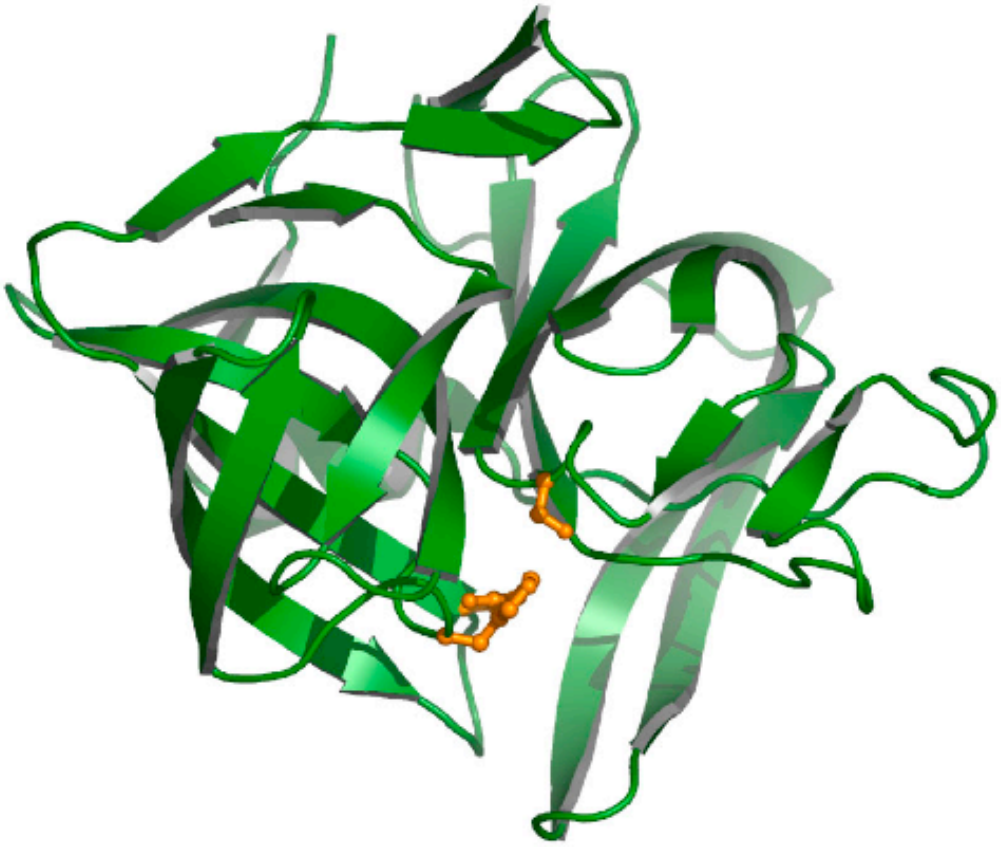


Figure 1.1: Structure of α-Lytic Protease (αLP).
αLP is a serine protease (catalytic triad in gold ball-and-stick) of the chymotrypsin family.

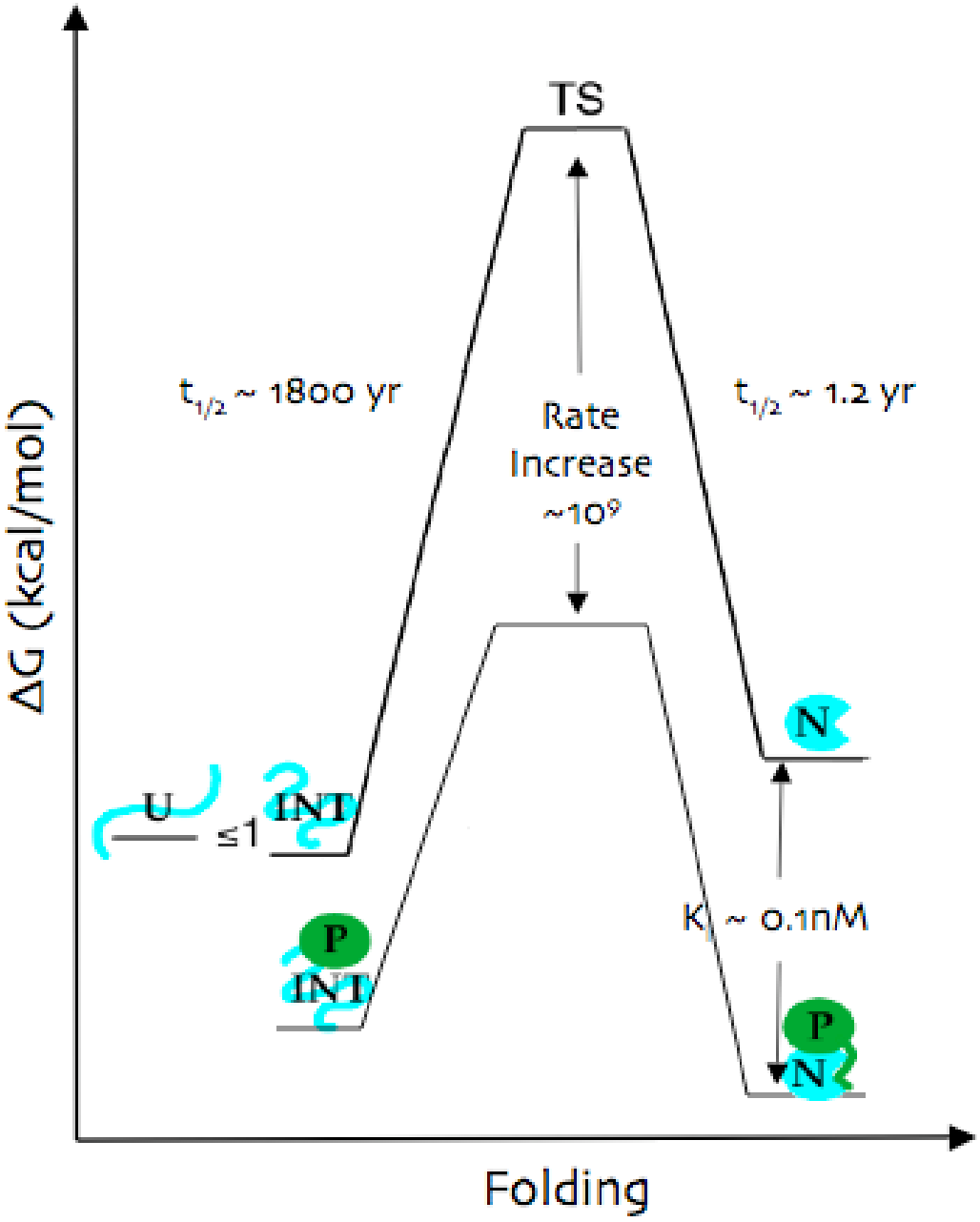


Figure 1.2: αLP folding free energy diagram.
αLP (blue chain or pacman) requires the Pro Region (green) to catalyze folding because the αLP folding free energy barrier is so high and the αLP native state is thermodynamically unstable. The Pro Region accelerates folding >10^9 and binds very tightly to the native state, causing efficient αLP folding. Upon removal of the Pro Region, the αLP Native state faces a large unfolding free energy barrier.

The enormous folding free energy barrier, combined with the thermodynamically unfavorable native state, explains why αLP requires Pro Region for its proper folding. To understand how the Pro Region performs this difficult task, careful kinetic studies of αLP refolding catalyzed by the Pro Region were undertaken. The Pro Region catalyzes αLP folding by accelerating folding by a factor of $>10^9$, thus ensuring that folding occurs in a biologically relevant timescale(Peters, Shiau et al. 1998).

To understand the mechanism of Pro-dependent folding in detail, the catalyzed folding reaction was studied *in vitro* using the Pro Region in *trans*(Peters, Shiau et al. 1998). The Pro region binds to Int through a β-sheet in the C-terminal domain of the Pro region to a unique β-hairpin in the C-terminal domain of the protease, forming a 5-stranded β-sheet(Peters, Shiau et al. 1998; Sauter, Mau et al. 1998). The Pro Region then binds incredibly tightly to the folding TS, causing the enormous acceleration in folding kinetics. It is clear that interactions throughout the both domains in the Pro Region are key for this catalysis, since mutation of residues from both the N- and C-domains severely diminish the catalytic efficiency of the Pro Region(Peters, Shiau et al. 1998; Cunningham, Mau et al. 2002).

The Native-αLP:Pro complex is also lower in free energy than the Int:Pro complex, which spontaneously drives the reaction toward to folded form of the protease(Peters, Shiau et al. 1998). Furthermore, the Pro Region is an extremely potent inhibitor of the protease, thus ensuring that the protease does not endanger the host cell through premature activation(Sohl, Shiau et al. 1997). Once the native protease is folded and secreted, the Pro Region can be proteolytically removed through cleavage of an unstructured loop by exogenous protease to prevent the Pro Region catalyzing the reverse

reaction of unfolding the protease(Cunningham and Agard 2004). These results highlight how the Pro Region has been highly optimized through evolution to catalyze the proper folding and function of αLP.

Costs and Benefits of Kinetic Stability

The Pro Region, which is necessarily degraded for active αLP production, is almost as large as αLP itself. What is the functional benefit of evolving a kinetically stable folding landscape that is worth such a cost?

αLP, which is secreted into soil(Fuhrmann 2003), normally exists in an extremely harsh environment and must have evolved to withstand fluctuating conditions and high protease concentrations. Kinetic stability seems to have been the mechanism evolved to withstand these pressures. First, the large unfolding free energy barrier severely reduces the sampling of the unfolded state(Sohl, Jaswal et al. 1998), which is highly protease-sensitive due to extensive exposure of flexible loops. However, a large unfolding barrier by itself is not sufficient to prevent proteolysis; partially unfolded forms(Bai, Sosnick et al. 1995) of the protease would render the protein susceptible to proteolytic destruction as well(Rupley 1967; Park and Marqusee 2004; Park and Marqusee 2005). However, local unfolding and breathing motions are greatly suppressed in αLP. Hydrogen exchange experiments showed that αLP has extremely high amide protection factors(Jaswal, Sohl et al. 2002), many higher than has been observed for any other protein(Huyghues-Despointes, Langhorst et al. 1999; Huyghues-Despointes, Scholtz et al. 1999). In addition, the highly protected amides are distributed throughout both domains rather than localized to a small core, as has been found for most other proteins(Li and Woodward

1999). Crystallographic B-factors, another measure of native state motion, are extremely low(Fujinaga, Delbaere et al. 1985; Rader and Agard 1997; Fuhrmann, Kelch et al. 2004), further illustrating the rigidity of the αLP native state.

This suppression of native state dynamics functionally results in greatly extended protease lifetime, even in the presence of highly proteolytic environments. Survival assays, in which αLP is incubated with the orthogonal proteases trypsin and chymotrypsin, have shown that αLP outlasts its thermodynamically-stable counterparts by a factor of 100(Jaswal, Sohl et al. 2002). αLP and the homolog *Streptomyces griseus* Protease B (SGPB) are degraded at a similar rate as the global unfolding rate(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004), while the thermodynamically stable trypsin is destroyed ~50-fold faster than it unfolds(Truhlar, Cunningham et al. 2004). Therefore, kinetic stability provides a mechanism by which protein lifetime can be greatly extended, through both extremely slow unfolding and tightly suppressed native state dynamics.

In addition, kinetic stability may provide more flexible paradigm for the adaptation of a protein to its environment. Because only the interactions that are broken during the N-TS transition are important for stability of the protein, there are considerably less interactions that need to be optimized to provide longevity in a certain environmental niche. This is in stark contrast to a thermodynamically stable protein, where all interactions are crucial to its overall stability. Furthermore, kinetically stable proteins may adapt to a wide variety of conditions through relocation of sensitive residues to regions unchanged in the N-TS transition. The concept of kinetic stability as

a flexible paradigm to rapidly and effectively evolve new behavior will be revisited in Chapter 4.

The advantages of kinetic stability do not come without a cost. First, there is the obvious cost of production of the large Pro region that gets destroyed for every protease molecule. This cost is non-trivial as the pre-proprotease total length can be about twice that of the actual protease. Furthermore, the tradeoff between the size of the Pro Region and the increase in kinetic stability does not appear to linear; relatively small increases in the unfolding barrier cause much larger increases in the folding barrier, thus necessitating a larger and more sophisticated Pro Region(Truhlar, Cunningham et al. 2004). Additionally, the greater kinetic stability translates into steeply decreased thermodynamic stability(Truhlar, Cunningham et al. 2004). Therefore, the Pro Region must bind extremely tightly to the native protease to drive the thermodynamics of folding towards production of the native state.

Physical and Structural Basis for Kinetic Stability

While it is understood why αLP and the other Pro-dependent proteases have evolved kinetic stability, the actual structural basis for this behavior is poorly understood. Understanding the structural basis for kinetic stability in the alpha-Lytic Protease class will be important for understanding many other proteins are kinetically stable, such as capsid protein SHP(Forrer, Chang et al. 2004), lipase,(Rodriguez-Larrea, Minning et al. 2006) pyrrolidone carboxyl peptidase,(Kaushik, Ogasahara et al. 2002) and subtilisin BPN',(Eder, Rheinnecker et al. 1993). Moreover, molecular evolution in some thermodynamically stable proteins, such as thioredoxin(Godoy-Ruiz, Ariza et al. 2006)

and p53(Butler and Loh 2005; Butler and Loh 2006), appears to be controlled by kinetic stability (i.e. unfolding rate). Additionally, the rate-limiting step in the formation of amyloid fibers, ordered aggregates of protein associated with many human diseases, is the rate of unfolding of the parent protein. (Canet, Sunde et al. 1999; Thirumalai, Klimov et al. 2003; Ohnishi and Takano 2004; Johnson, Wiseman et al. 2005) Thus, the basic concepts that underlie kinetic stability will be valid for most, if not all, proteins as well as for understanding the basis for many diseases. While much is still unknown, there are some clues as to the structural mechanisms by which these landscapes can be modulated.

First, a quasi-thermodynamic analysis of the unfolding free energy barriers for αLP and SGPB have been conducted, revealing rich insight into the physical chemistry of the folding landscapes for these proteases(Jaswal, Truhlar et al. 2005). By measuring the temperature dependence of unfolding for αLP and SGPB and folding for αLP, the $\Delta C_p^{\ddagger}$, $\Delta H^{\ddagger}$, and $\Delta S^{\ddagger}$ terms for these reactions were determined through Eyring analysis(Jaswal, Truhlar et al. 2005). These studies showed that the free energy barrier for folding was entropic in origin(Jaswal, Truhlar et al. 2005). Likewise, calorimetric data showed that the Int is stabilized over N through entropy(Sohl, Jaswal et al. 1998). The physical basis for these results could be explained that αLP, and all its Pro-dependent homologs, have abnormally high glycine content(Jaswal 2000). While the mammalian digestive proteases and other members of the trypsin family that fold autonomously are 7-11% glycine, the Pro-dependent proteases in the αLP subfamily are 16-20% glycine. The increased flexibility afforded by the extra glycine residues is expected to stabilize the αLP Int over N by ~ 7 kcal/mol(D'Aquino, Gomez et al. 1996), which could partially explain the observed metastability. The TS is expected to be

considerably more ordered than Int as well, so high glycine content could partially account for the extremely slow folding rate.

It is not surprising that the folding reaction, in terms of both the free energy barrier and thermodynamics, is entropically unfavorable(Jaswal, Truhlar et al. 2005) because a folding reaction will necessarily lose configurational entropy as the polypeptide chain becomes more ordered. However, it was truly unexpected that the unfolding free energy barrier for both αLP and SGPB is entropic in origin as well(Jaswal, Truhlar et al. 2005). Normally, entropy favors formation of the TS at physiological temperature due to the increased configurational entropy of this more plastic state. The entropic barrier to unfolding in αLP and SGPB indicates that the system becomes more ordered in the TS. Because N is already completely ordered, there must be significant ordering of solvent in the TS, yielding deep insight into the physical characteristics of the TS.

Additionally, the $\Delta C_p^{\ddagger}$ and m-values (from the temperature and denaturant dependence of unfolding rate, respectively) have provided independent estimates for the nativeness of the TS for αLP and SGPB. The fractional $\Delta C_p^{\ddagger}$ and m-values report on the fraction of non-polar and total surface area exposed in the TS, respectively(Myers, Pace et al. 1995). The αLP fractional $\Delta C_p^{\ddagger}$ and m-values indicate that the TS is highly native-like (0.62 and 0.79, respectively), although the results for SGPB were not as clear (0.31 and 0.93, respectively.)(Truhlar 2004) While the cause of this large discrepancy between the fractional m-value and $\Delta C_p^{\ddagger}$ in SGPB is unknown, it could be the result of TS movement with temperature(Privalov and Makhatadze 1990; Dalby, Oliveberg et al. 1998), disproportionate exposure of nonpolar surface area(Main, Fulton et al. 1999),

and/or differences in the viscosity(Plaxco, Simons et al. 1998; Jacob and Schmid 1999; Perl, Jacob et al. 2002). Regardless, the αLP results clearly indicate that the TS is quite native-like for this protein.

Taken together, the quasi-thermodynamics parameters of unfolding dictates a TS structure in which there is significant ordering of solvent without much loss of native structure. To account for this, Sheila Jaswal postulated a structural model for the αLP TS in which the N- and C-domains separate from each other, but remain relatively intact themselves(Jaswal 2000) (Figure 1.3). Solvent would then be significantly ordered in the newly formed crevice at the domain interface(Jaswal 2000). Although this model is consistent with the thermodynamic data, actual physical evidence for this mode of unfolding is lacking. Experiments providing the first physical evidence for this model are described in Chapters 4 and 6.

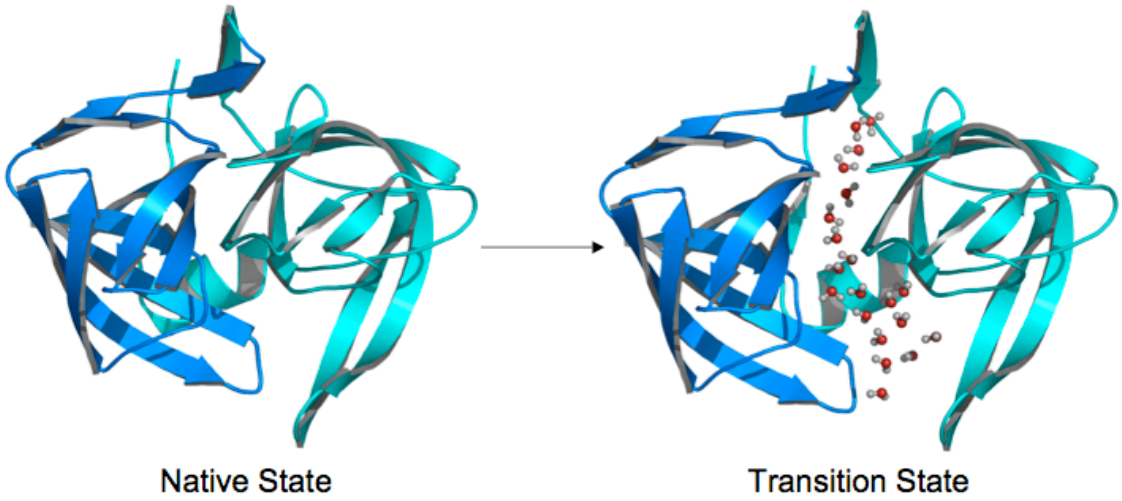


Figure 1.3: Proposed Unfolding TS Model for αLP.
The N- and C-domains themselves remain relatively intact but separate from each other, allowing solvent to enter into the inter-domain crevice. This model was proposed based purely on the thermodynamics of unfolding, with no physical evidence supporting it. One of the goals of this work is to test this model.

The quasi-thermodynamic analysis of αLP and SGPB unfolding also showed that the $\Delta C_p^{\ddagger}$ for these proteins are quite large as compared to a set of thermodynamically stable proteins(Jaswal, Truhlar et al. 2005). This implies that there is a significant network of native interactions that are broken only upon formation of the TS(Jaswal, Truhlar et al. 2005), which could explain the ultra-cooperative unfolding transition observed for these proteins(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004)(see above). In support of this conclusion, it was recently found that the N- and C-domains folding in an interdependent manner(Cunningham and Agard 2003), in contrast to trypsin and chymotrypsin, in which there is independent folding fo the N- and C-domains(Higaki and Light 1986; Light and al-Obeidi 1991). The αLP N- and C-domains can rapidly dock onto each other, but the rate-limiting step for folding is a cooperative rearrangement to form N(Cunningham and Agard 2003), thus suggesting that the domain interface plays an important role in αLP folding.

While these studies have provided general insight into the source of kinetic stability, they lack information about the role of specific interactions. Some progress towards this goal was made by screening for αLP mutants that fold faster than wild-type. After screening through >3 $x10^5$ colonies, the same mutant was recovered: R138H, G183S(Derman and Agard 2000). Remarkably, this mutant affects the uncatalyzed and Pro-catalyzed folding reactions similarly, suggesting that the TS structure for these disparate reactions are similar despite the 10^9 difference in rate(Derman and Agard 2000). However, exactly how this mutant exerts its effect remains to be determined, but will certainly provide important information into how kinetic stability is modulated.

The large differences in the folding landscapes for αLP and SGPB have been examined to discern structural features that modulate kinetic stability. A comparison of small and large Pro region dependent protease sequences revealed that there are residues whose identity covary according to Pro size (see Chapter 2). In SGPB and other small pro proteases(James, Sielecki et al. 1980; Nienaber, Breddam et al. 1993; Kitadokoro, Tsuzuki et al. 1994; Huang, Lu et al. 1995), there is a type I' or II' turn in the C-terminal β-hairpin while αLP contains a less energetically favorable type I turn(Fuhrmann, Kelch et al. 2004). To test whether this structural element could be causing differential folding behavior for these proteases, a chimeric αLP protein was made in which the αLP turn was substituted with that of SGPB(Truhlar and Agard 2005). Although not conclusive, the results were consistent with the importance of the turn packing with the N-domain in the unfolding process, and with the model of the unfolding Transition State that was previously proposed(Jaswal 2000) (Figure 1.3). Further experimentation, including use of Molecular Dynamics simulations of unfolding(Daggett 2006), will be necessary to discern the true role of this structural element in the αLP folding landscape.

Regardless, the strategy of homolog comparisons combined with chimeric variants holds much promise for investigation of the structural basis for kinetic stability and will be seen again in Chapters 3, 4, 5, and 6. In fact, another possible putative key region to kinetic stability was identified through a comparison of SGPB and αLP. The identities of several residues clustered in the core covaried according to the size of the corresponding pro region(Fuhrmann, Kelch et al. 2004). This cluster completely surrounded a single residue that is absolutely conserved as a planar residue (Phe, Tyr or, rarely, His). Despite the sequence conservation of this residue, its conformation covaries

according to Pro Region size, as well. In the large Pro Region-containing protease αLP, a 0.83Å crystal structure unambiguously shows that this residue is distorted from planarity by a remarkable ~6º, due to extremely tight packing interactions with the covarying residues(Fuhrmann, Kelch et al. 2004). It has been postulated that this distortion could be playing a key role in determining the kinetically stable landscape of αLP through modulation of free energy barrier height and/or protein dynamics(Fuhrmann, Kelch et al. 2004). Further exploration of the role of the geometric strain energy in determining the energy landscapes for these proteases is described in Chapter 3.

Specific Aims

1) What are the evolutionary benefits of kinetic stability?
2) How can kinetic stability be utilized to adapt proteins to extreme conditions, such as high temperature and/or acidity?
3) What are the physical origins of kinetic stability?
4) What is the structure of the Transition State?
5) What is the role of geometric strain in kinetic stability?

Chapter 2: Phylogenetic analysis of the alpha-Lytic Protease class of Pro-dependent Proteases

Preface

The structural mechanisms underlying kinetic stability are unknown. Because of the extremely slow kinetics of folding and unfolding, classical folding techniques such as comprehensive Phi-value analysis are not reasonable. To identify the structural determinants for this behavior, we used evolutionary relationships to gain a foothold into the sequence determinants that modulate kinetic stability.

Dr. Luke Rice and I co-wrote the MatLab program to perform the MSA-PCA analysis. I am responsible for all other analyses. Some of the results from the covariation analysis were published in Fuhrmann et al (2004) *Journal of Molecular Biology* 338, 999-1013, for which I was second author.

Introduction

Alpha-Lytic Protease and its homologs have been useful model systems for understanding protein folding. The stability of these proteins is under kinetic control(Baker, Sohl et al. 1992), so these proteases have been extremely useful in the study of folding barriers. Because the concepts that underlie kinetic stability will advance our understanding of folding mechanisms in general, insight into the structural basis for this behavior will be of great use to the general folding community. However, a detailed description of the mechanism of folding landscape has yet to be determined.

To understand the molecular determinants of kinetic stability, insight into the structural attributes of the Transition State are necessary. The common means of examining folding transition state structure is phi-value analysis, which can measure the contribution of a sidechain to the folding Transition State on a residue by residue basis (Matouschek, Kellis et al. 1989). However, the extremely slow kinetics of folding for these proteases makes full scale phi value analysis experimentally intractable. Therefore, other means are necessary to elucidate the mechanisms by which kinetic stability is modulated.

Although conventional methods are unfeasible, it is possible to obtain insight into the determinants of kinetic stability using evolutionary relationships between homologs with altered folding landscapes. There are significant differences in the folding landscapes of proteases with different sized pro regions(Truhlar 2004). Large pro protease αLP has larger barriers to unfolding and folding, thus necessitating evolution of larger pro region(Truhlar 2004). It is assumed that this trend extends over other family members. Thus, a large-scale comparison of small and larger Pro region-dependent proteases could elucidate the mechanisms by which the degree of kinetic stability can be modulated. By examining the sequence variation throughout and between these two sub-classes, it may yield insight into the specific residues and interactions that are important for the alteration of their folding landscapes.

Results and Discussion

As a first step towards this goal, we collected sequences of 71 homologs in the NCBI database by performing a BLAST search using the αLP protease domain as the

query (Table 1). A majority of the proteases were from eubacteria, with only one of the retrieved sequences represented in archaeabacteria (from *Pyrococcus furiosus*), and three in eukaryotes (*Phaeosphaeria nodorum, Metarhizium anisopliae,* and *Arabidopsis thaliana*). Soil organisms are highly represented in the set of proteases, which may represent a proclivity of these organisms to utilize these long-lasting enzymes for catabolic purposes, or could be a result of the extensive sequencing of soil bacteria genomes. Three proteases were identified in thermophilic or extreme thermophilic organisms, *Thermobifida fusca* (see Chapters 6 and 7), *Pyrococcus furiosus* (see brief note in Chapter 9) and *Corynebacterium efficiens*. Analysis of these proteases could elucidate structural determinants of thermophilic behavior and the modulation of kinetic stability at high temperature (see Chapters 6 and 7.)

To categorize the proteases according to their pro region size, the sequence of the preproprotease was analyzed to determine the length of the signal sequence (SS), Pro Region, and protease domains. The SignalP server(Bendtsen, Nielsen et al. 2004) was used to estimate the SS, to high accuracy. In almost all sequences, the SS could be determined with >99% estimated accuracy. The pro region length was then estimated from the end of the signal sequence to the beginning of the protease domain, which was determined based on homology to the closest relative for which the protease domain is known. The pro regions fell into three categories of length: small (50-98 residues), medium (119-146 residues) or large (151-198 residues). In all there are 9 small, 18 medium, and 36 large proteases in the set of proteases. Some proteases had additional C-terminal domains, most often of ricin domains that are known to bind carbohydrates.

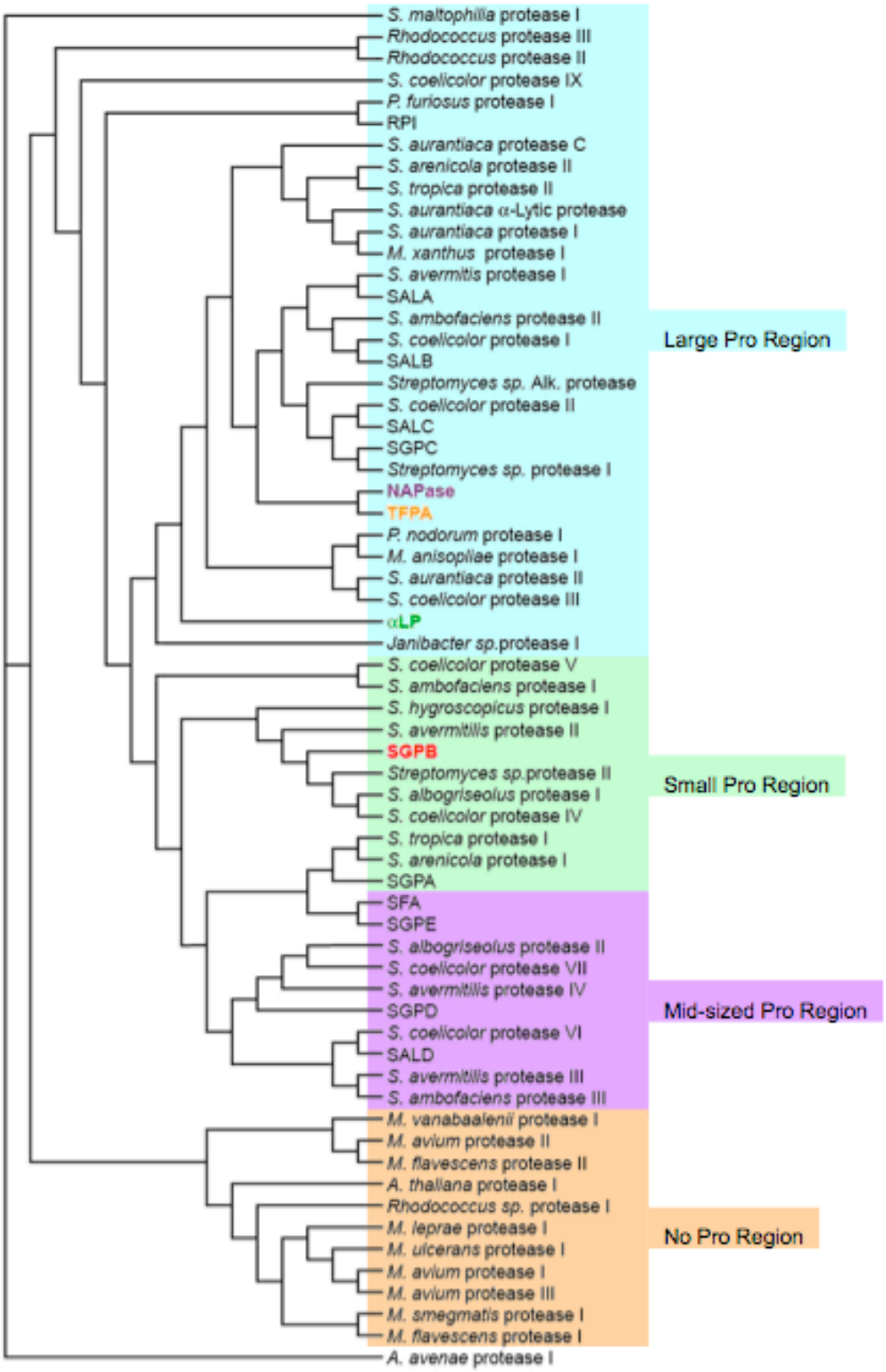


Figure 2.1: Phylogenetic Tree of the Pro-dependent proteases and close relatives based on a multiple sequence alignment (MSA).

Figure 2.1: Phylogenetic Tree of the Pro-dependent proteases and close relatives based on a multiple sequence alignment (MSA). (see prior page)
The proteases are in different clades that correspond to the size or presence of their associated Pro domains, even though only the protease domains were used in the MSA. Note that the branches can be inverted along any node without a change in tree topology.

To understand evolutionary relationships and gain into the functional properties of these proteases, a phylogenetic analysis was performed. As a first step, a Multiple Sequence Alignment of αLP homologs was created using ClustalW(Chenna, Sugawara et al. 2003). A phylogenetic tree was then created using only the protease domains, using the Neighbor Joining method(Saitou and Nei 1987; Bendtsen, Nielsen et al. 2004). Quite remarkably, the proteases can be separated by the size of their respective pro regions despite the fact that this domain was not actually used in the generation of the tree (Figure 2.1). This means that the information about pro domain size, and therefore the degree of kinetic stability(Truhlar 2004), is encoded in the sequence of the protease domain itself. Thus, it may be possible to extract structural insight into kinetic stability directly from a sequence comparison of large and small pro-dependent proteases.

A more sophisticated method than standard phylogenetic tree was employed to further examine the evolutionary relatiosnships between different proteases. A matrix of pair-wise (protease to protease) sequence differences was built from a multiple sequence alignment in ClustalW(Chenna, Sugawara et al. 2003). To illustrate the total variation in these sequences, Principal Components Analysis (PCA) was performed on the matrix using MatLab and the top two eigenvectors were plotted (Figure 2.2). The distance between points signifies the degree of sequence variation between captured by these two eigenvectors, which account for ~10% and ~50% of the total variation in the data set, respectively. Quite remarkably, the second eigenvector separates the proteases into two

major classes, according to their pro size. Significantly, this result illustrates that the sequence variation due to pro region size accounts for, on average, ~10% of the sequence variation of each protease.

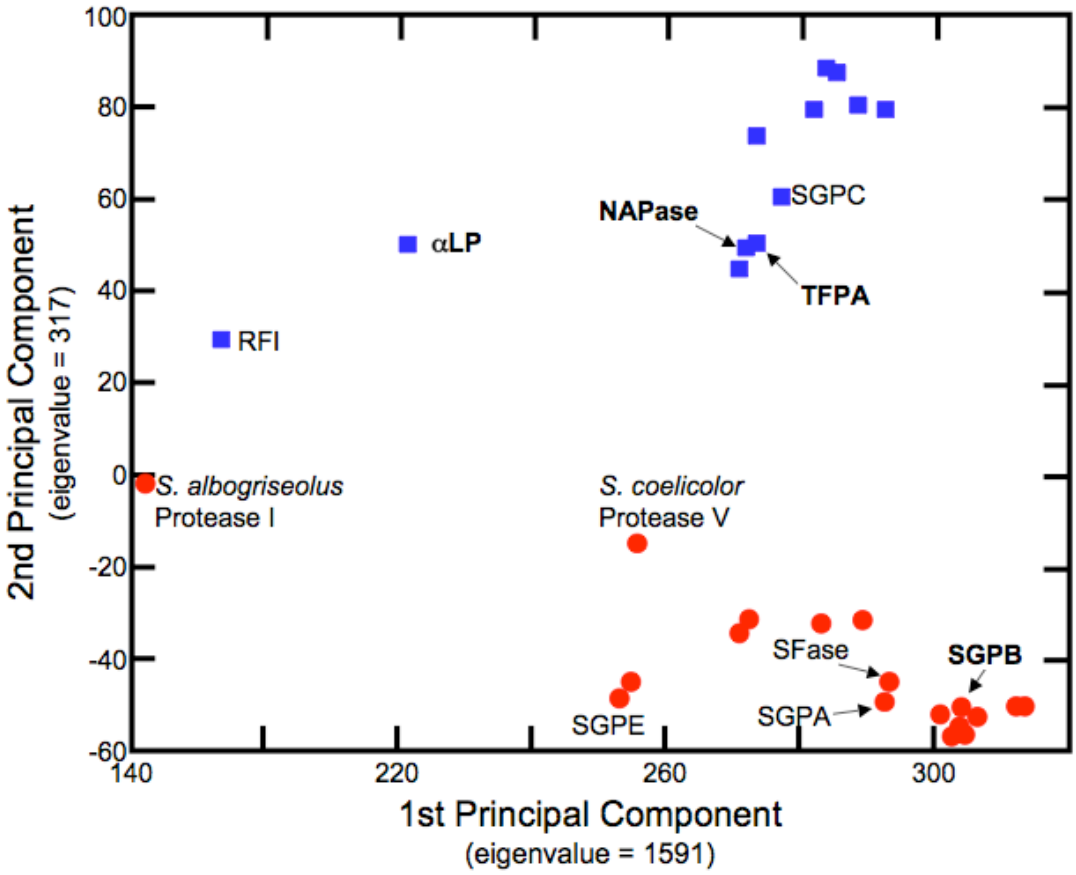


Figure 2.2: Multiple Sequence Alignment Principal Component Analysis (MSA-PCA) of the Pro-dependent proteases.
Proteases utilizing a large pro region (blue squares) are separated from those utilizing a medium or small pro region along the second principal component. αLP is an outlier along the first principal component. Only the protease domains were present in the MSA.

Clearly, there is important information as to the structural mechanisms of modulating folding landscapes encoded in the observed sequence variation. To elucidate this relationship, we investigated sequence covariation on a residue-by-residue level.

Specifically, we were looking for residues that were conserved as one type of amino acid in the small Pro-dependent prtoeases, but conserved as another in the large Pro-dependent proteases. To simplify the analysis, the proteases utilizing a medium sized Pro Regions were classified as small Pro Region dependent proteases. This is not unreasonable as these two versions of proteases form one clade on the phylogenetic tree (Figure 2.1).

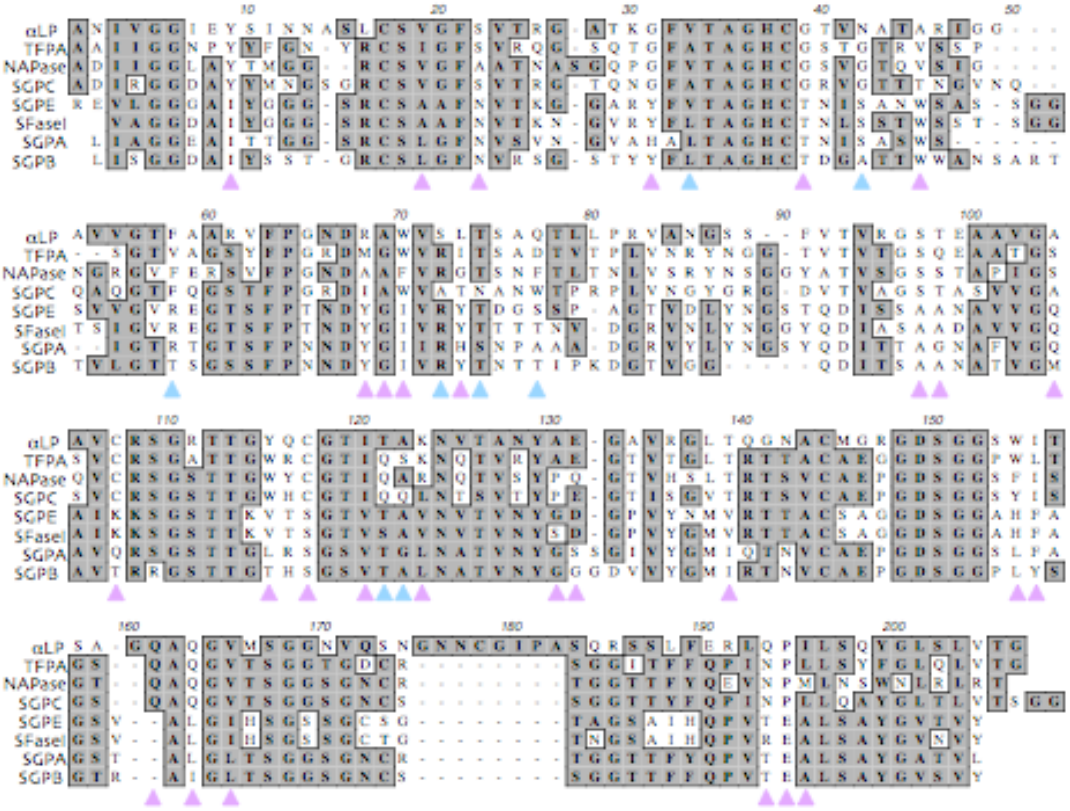


Figure 2.3: Multiple Sequence Alignment (MSA) illustrating the position of covarying residues.
Residues identified with a violet triangle are strongly covarying while those with a blue triangle covary weakly. An MSA of 35 proteases was used to identify the covarying residues, while only 8 proteases are shown here for ease of visualization: αLP, TFPA, NAPase and SGPC (large pro region), SGPE and SFase1 (medium pro region), SGPA and SGPB (small pro region.)

Many strongly covarying residues were identified throughout the entire protein as shown in Figures 2.3. The structural mapping of covarying residues (Figure 2.4) is quite remarkable as these residues form a continues network from the N- to C-domains, very reminiscent of the allosteric networks identified by the MSA-Statistical Coupling analysis pioneered by the Ranganathan group(Lockless and Ranganathan 1999; Suel, Lockless et al. 2003; Socolich, Lockless et al. 2005). This is not entirely surprising since both approaches share many similarities, in that they probe for sequence changes at a residue-by-residue level in response to some external perturbation (in the Statistical Coupling method, the identity of another residue; in our method, the size of the associated pro region.)

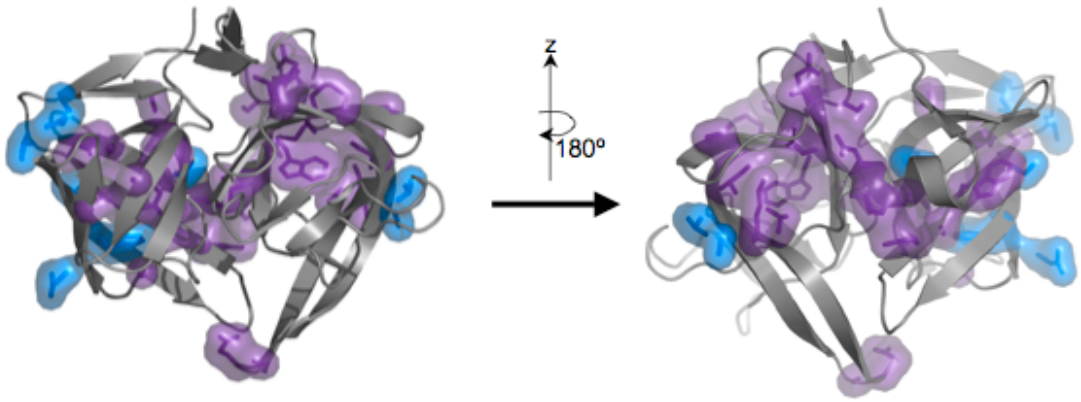


Figure 2.4: Mapping covariation onto the αLP structure.
Residues identified as covarying (figure 3) are displayed with Van der Waals surfaces colored violet (strongly covarying) or blue (weakly covarying.)

More covarying residues were found in the C-domain (21 versus 17). This is not unexpected considering that the Pro Region binds to the C-domain almost exclusively. However, it was surprising to discover that the surface of the protease upon which the Pro

Region N-domain binds is largely devoid of covarying residues. Since the Pro N-domain is only present in the large Pro Regions, it was expected that there would be very large covariation along this interface. The residues along this interface are not conserved uniquely amongst the large Pro region proteases, suggesting that the interactions between the protease C-domain and Pro N-domain are highly variable. This is supported by a crystal structure of the Pro:αLP complex, where this interface consists mainly of water-mediated hydrogen bonds(Sauter, Mau et al. 1998). Despite this plasticity, these interactions are actually very important for catalysis, as mutations in the Pro N-domain are required for efficient (Cunningham, Mau et al. 2002). The low specificity of the surface interactions between the protease C-domain and the Pro N-domain suggests that the residues that control folding and unfolding in the isolated protease are present elsewhere.

Remarkably, a majority of the covariation resides in the cores of the protease (Figure 2.4). For example, the core of the N-domain displays remarkable covariation, with many of the residues pointing into the β-barrel strongly covarying and weaker covariation on the periphery. Since the cores of proteins provide most of the folding energy, this observation could shed some much-needed light into the modulation of kinetic stability. However, it is not clear from a cursory structural analysis how this covariation could yield the observed differences in folding behavior. Further investigation of the covariation in this region would definitely be interesting.

Other regions of the structure display remarkable covariation between the 2 clades of proteases, especially a large grouping of residues in the core of the C-domain. This cluster surrounds residue 228, which is conserved as Phe, Tyr, or His throughout all the

Pro-dependent proteases and most of the chymotrypsin family. Interesting, the covariation around residue 228 causes its sidechain to adopt two completely different rotamers depending on Pro region size. In the proteases with a small Pro Region or those without one, the aromatic ring of the sidechain is pointed toward the active site. However, this orientation is prohibited in the large Pro region-dependent proteases due to the presence of a large hydrophobic residue at position 199 which sterically occludes the 228 sidechain. Instead, the ring is rotated by ~120 degrees around χ_1 and is sandwiched between residues 181 (usually a Thr) and 199 (usually aromatic), while other residues in contact or in the vicinity of residue 228 covary to afford this large rearrangement in the hydrophobic core.

Quite remarkably, a 0.83Å resolution crystal structure of αLP unambiguously revealed that Phe228 is so tightly packed between Thr181 and Trp199 that the phenyl ring is distorted by 5.8º from planarity(Fuhrmann, Kelch et al. 2004). This distortion is distributed throughout the ring as all bonds in the sidechain are distorted from normal. This energetic penalty for creation of this distortion has been estimated to be 4.2 kcal/mol by quantum mechanical calculations(Fuhrmann, Kelch et al. 2004), which is approximately the same difference in energy between the Native and Intermediate states(Sohl, Jaswal et al. 1998) and overall stability difference between αLP and SGPB(Truhlar, Cunningham et al. 2004). Thus, this distortion could possibly account for the large differences in kinetic stability between αLP and SGPB. This hypothesis is tested in Chapter 3.

Some varying regions have been previously tested for their effect on the folding landscape. One such region is at the turn of a unique β-hairpin in the C-domain. The

sequence in the small Pro-dependent proteases causes a type I' or II' turn(James, Sielecki et al. 1980; Nienaber, Breddam et al. 1993; Kitadokoro, Tsuzuki et al. 1994; Huang, Lu et al. 1995) while those with a large pro region have a type I turn(Fuhrmann, Kelch et al. 2004). Since this β-hairpin is unique to the pro-dependent proteases and is known to be instrumental in folding catalysis(Peters, Shiau et al. 1998), it was suspected that the more favorable type I' turn in the small Pro proteases induces faster protein folding. To test this hypothesis, a chimeric variant was made in which the β-turn sequence from SGPB was substituted into αLP(Truhlar and Agard 2005). However, there was no effect on the uncatalyzed folding reaction and only a small, 3-fold acceleration of the pro-mediated folding reaction, suggesting that this hypothesis is not correct. This is not entirely unsurprising since the timescale of β-hairpin rearrangements is on the order of microseconds(Munoz, Thompson et al. 1997), while that of protease folding is on the order of days for SGPB(Truhlar, Cunningham et al. 2004) to millennia for αLP(Sohl, Jaswal et al. 1998). A small, 2-fold deceleration of unfolding is consistent with this β-turn playing some role in the unfolding reaction(Truhlar and Agard 2005), although it is difficult to tell whether this small effect is significant. More experiments are necessary to discern the true role of this structural element in modulating the kinetically stable energy landscape of these proteases.

In addition, the large Pro region proteases contain a disulfide bond between residues 137 and 159 of the protease C-domain, while the small Pro proteases do not. This disulfide is the only interaction that lies on the surface of the C-domain that is in contact with the Pro N-domain, suggesting that it may play some role in determining the degree of kinetic stability and/or assist in the catalysis of folding. However, the disulfide

bond, by greatly lowering the entropy of the unfolded state, would be expected to accelerate folding rather than decelerate as would be predicted. To address the role of this disulfide bond in kinetic stability, mutants of αLP were made in which this disulfide was removed by mutation of residues 137 and 159 to Ala and/or Ser in various combinations(Truhlar 2004), but none of the variants expressed to detectable levels. Creation of a new disulfide in the context of SGPB could be advantageous in identifying the role of this disulfide in the folding landscapes for these proteins.

There are many other sites throughout the protein in which residues covary according to Pro Region size that haven't been fully explored. Some of these residues are scattered throughout the structure, and there isn't a clear model for how they could be modulating the folding landscapes of these proteases. While mutation of these residues may yield significant insights, these experiments need more structural and theoretical background to guide their implementation. A more sophisticated strategy for examining covariation could greatly assist in this realm. By performing PCA on the covariation of amino acid identities residue-by-residue, a more complete and quantified assessment of covariation could be attained. To implement this approach, a matrix of all-protease-against-all comparisons at each residue would be built, using a substitution matrix, such as BLOSUM(Henikoff and Henikoff 1992), as a scoring function. Subsequent clustering could group residues according to their pattern of covariation to discern which residues have coupled covariation and could lead to an understanding of how energetic coupling throughout the structure is obtained in a quantitative manner analogous to the Statistical Coupling technique of Ranganathan and coworkers(Lockless and Ranganathan 1999; Suel, Lockless et al. 2003; Socolich, Lockless et al. 2005).

Materials and Methods

Protease sequences were obtained by using BLAST(Altschul, Gish et al. 1990) of the NCBI sequence database (http://www.ncbi.nlm.nih.gov/BLAST/) with the αLP protease sequence as the query. The signal sequence was determined with the SignalP server(Bendtsen, Nielsen et al. 2004), and the Pro sequence was estimated based off of homology with the closest known relative.

MSAs were created using ClustalW(Chenna, Sugawara et al. 2003) with the BLOSUM substitution matrix(Henikoff and Henikoff 1992). Phylogenetic trees were made using the Neighbor Joining mehtod(Saitou and Nei 1987) in MacVector. MSA-PSA was performed using the pairwise scores output in ClustalW from the MSA as values in the diagonalized matrix. This was accomplished using a program co-written with Dr. Luke Rice in MATLAB.

Covariation was estimated using complete MSAs of the Pro-dependent proteases. Covariation in this case is defined as differential conservation at each position dependent on the size of the corresponding pro region. In other words, the position could be conserved as one type of amino acid in the small pro proteases, but unconserved in the large pro proteases, or vice versa. Alternatively, the residue could be conserved within each class, but not conserved across the classes. Each position was visually inspected for covariation and scored accordingly.

Chapter 3: Sidechain Distortion as a Novel Mechanism for Functional Modulation of Folding Landscapes

Preface

A 0.83Å resolution structure of αLP showed that a buried Phe sidechain was distorted from its planar geometry by 5.8º. Based on a pattern of covariation (see Chapter 2), this distortion was proposed to contribute to the kinetic stability of αLP. Here we show that this distortion does indeed contribute to αLP's kinetic stability, as well as other homologs. We propose that sidechain distortion is a relatively common mechanism for modulation of protein energetics for biological function.

This chapter is being written as a paper for submission as a letter to Nature.

First Paragraph

Proteins utilize various non-covalent interactions, such as Van der Waals forces, ionic and hydrogen bonds to perform necessary functions. Recent access to ultra-high resolution (<1.0Å) protein structures have revealed unprecedented detail of atomic geometry, such as the unexpected deformation of aromatic rings from planarity(Fuhrmann, Kelch et al. 2004; Kang, Devedjiev et al. 2004) (Figure 1). However, it is unclear what role, if any, this energetically-costly strain plays in protein function. The ultra-high resolution structure of alpha-Lytic Protease (αLP) indicated that residues surrounding a conserved Phe sidechain dictate a rotamer choice that leads to an ~6° distortion (Fig 1b) (Fuhrmann, Kelch et al. 2004). By contrast, in a closely related

protease, *Streptomyces griseus* Protease B (SGPB), the equivalent Phe adopts a different rotamer and is not distorted. Phylogenetic analysis of the αLP family of extracellular serine proteases reveals a remarkable bifurcation into those sequences that would be expected to induce distortion and those that would not (Fig 1C). This differential conservation suggests that the rotamer choice and, by implication, the Phe distortion plays a biologically relevant role. Here we report that the αLP Phe sidechain distortion is both conserved and functional. Atomic resolution crystal structures and kinetic studies of mutant proteins and family members provide compelling evidence that strain alters the folding free energy landscape by destabilizing the transition state (TS) compared to the native state (N). While sidechain distortion comes at a cost of foldability, it has the benefit of suppressing unfolding, thereby enhancing kinetic stability and increasing protein longevity under harsh extracellular conditions.

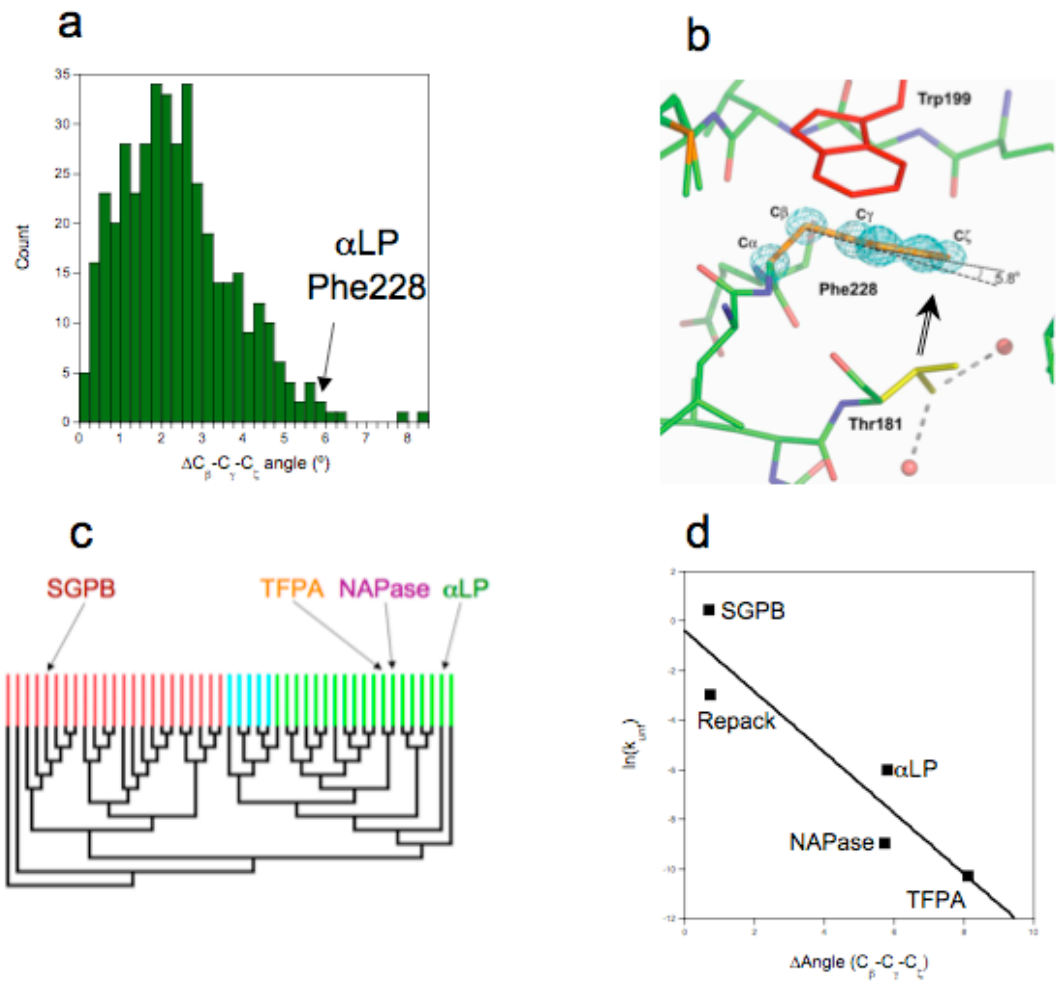


Figure 3.1: Distortion of phenylalanine sidechains
A) Histogram of Phe distortions present in ultra-high resolution (≤0.99Å) structures in the PDB. The distortion of αLP Phe228 is nearly 3 standard deviations from the mean. B) Distortion of Phe228 of αLP is caused by close contact of Thr181 (from a 0.83Å resolution structure(Fuhrmann, Kelch et al. 2004)). C) Phylogenetic analysis shows an evolutionary bifurcation in C-domain packing. Proteases expected to have a packing arrangement similar to SGPB or αLP are shown in red and green, respectively. Proteases exhibiting sequences that are a hybrid of the SGPB and αLP are shown in cyan. D) The unfolding barrier of the kinetically stable proteases is dependent on the distortion of residue 228. There is a strong correlation with the height of the unfolding barrier with the degree of Phe228 distortion for all homologs that have been tested, as well as the Repack mutant.

Body

The folding landscape of αLP is very different than typical, thermodynamically stable proteins in that its folding barrier is enormous ($t_{1/2}$ ~1800years) and the native state is less stable thermodynamically than the unfolded forms by an unprecedented 4 kcal/mol(Sohl, Jaswal et al. 1998). Therefore, unlike most proteins, αLP's stability is dictated solely by its extremely high unfolding barrier ($t_{1/2}$ ~1.2 year for N to TS) rather than the overall energetic difference between its native and unfolded states(Sohl, Jaswal et al. 1998). In αLP and other family members, kinetic stability is coupled to reduced native state dynamics resulting in enhanced longevity in harsh environments (Jaswal, Sohl et al. 2002; Truhlar 2004)(Kelch et al, in press). Although functionally very important, little is known about the structural basis for kinetic stability. Because SGPB has reduced kinetic barriers compared to αLP, analysis of sequence variation patterns and structural differences between the two proteases could identify regions of the molecule responsible for the observed functional differences. A prime candidate region is the differential packing in the C-terminal domain responsible for the distortion of Phe228 in αLP, suggesting that there might be a correlation between distortion and barrier height. Indeed, comparison of unfolding barrier height with distortion angle reveals a remarkable correlation across multiple homologs (Figure 1D, Figure S1), strongly suggesting the energetic and functional relevance of sidechain distortion.

To test this hypothesis, we created mutants that modify the interactions with Cβ and Cγ of Thr181 that induce the distortion (T181A; T181G) and examined their differential effects on folding and unfolding kinetics. (While the F228A mutation obviously removes the distortion, it creates a large cavity in the hydrophobic core which

severely complicates any interpretation of the folding data.) Atomic resolution structures (1.1Å; Fig 2A) demonstrate a mild reduction in strain (C_β-C_γ-C_ζ = 4.5º) in T181A and an almost complete elimination of strain in T181G (C_β-C_γ-C_ζ = 2.3º). This result is significant as it provides a more precise means to differentiate the folding impact of strain removal (T181A –> T181G) from the contributions of Thr Oγ1 and Cγ2 (wt -> T181A).

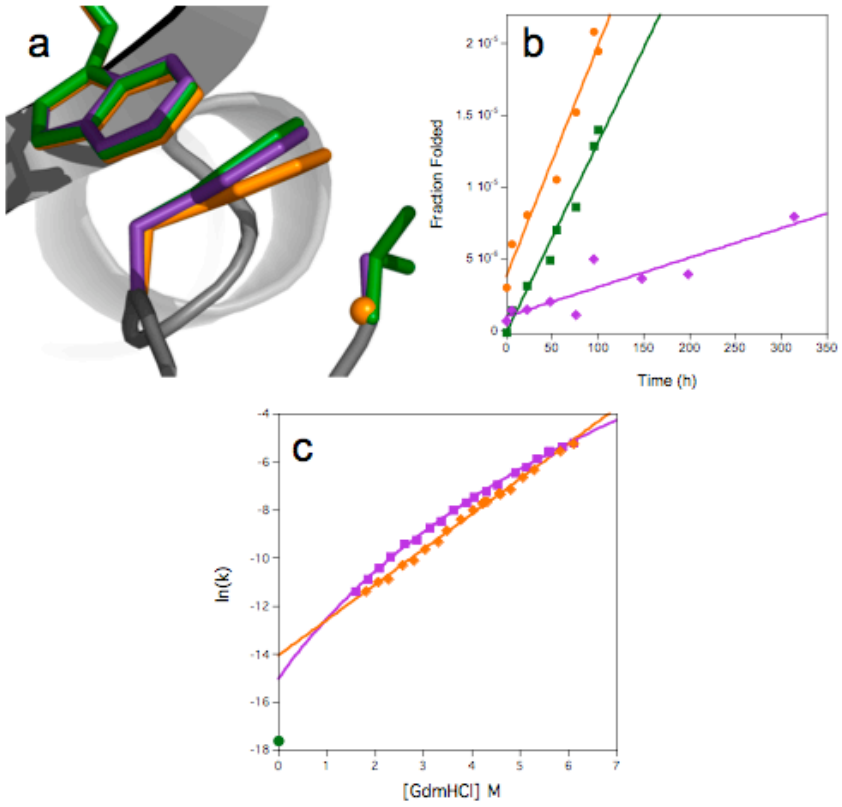


Figure 3.2: Structure and folding kinetics of Thr181 mutants.
A) Structures of wild-type αLP(Fuhrmann, Kelch et al. 2004) (green), T181A (1.08Å; violet), and T181G (1.10Å; orange). Phe228 distortion is still present in T181A, but absent in T181G. B) Time course of folding for WT-αLP (green squares), T181A (violet diamonds), and T181G (orange circles). T181A folds ~10 times slower than WT and T181G C) Extrapolated unfolding rates for T181A (violet) and T181G (orange), compared to the unfolding rate for WT (green). T181A and T181G unfold 35- and 13-times faster than WT.

To determine the energetic consequences of these mutations, folding rates were determined by measuring the very small amounts of active protease that fold within a practical time frame while denaturant titration coupled with tryptophan fluorescence provides an accurate measure of unfolding rates (see methods; (Sohl, Jaswal et al. 1998)). The rate of folding for the T181A mutant was ~7-fold slower than that of wild-type (Figure 2B) Remarkably, removing the entire sidechain at this residue (T181G) restored the folding rate to wild-type levels (Figure 2B). Both mutants unfold significantly faster than wild-type (T181A ~13-fold, T181G ~35-fold; Figure 2C). These folding and unfolding rates were converted to $\Delta\Delta G^{\ddagger}$ using standard Transition State theory to generate comparative free energy diagrams (Supplemental Figure 2). The removal of the C_{γ} and hydroxyl groups from Thr181 (T181A) causes destabilization of TS (1.0 kcal/mol) and N (2.4 kcal/mol). Complete removal of the sidechain and, therefore, distortion of Phe228 in the T181G mutant mitigates this destabilizing effect, both on TS and N (1.1 and 0.6 kcal/mol stabilization relative to T181A, respectively). The relative contribution of strain to kinetic vs thermodynamic stability can be quantified by Phi value analysis (Matouschek, Kellis et al. 1989). While Phi values typically range from 0 (no contribution to the TS) to 1.0 (fully contributing to the TS), the Phi value for strain (T181A->T181G) is a remarkable 1.9 (±0.5). We suggest that such non-canonical Phi-values will be hallmark for interactions that contribute significantly to kinetic stability. This confirms three key features of our hypothesis: 1) strain from the Phe228 distortion is indeed present in N, 2) this strain manifests itself in the TS as well, and 3) the strain in TS appears to be greater than in N.

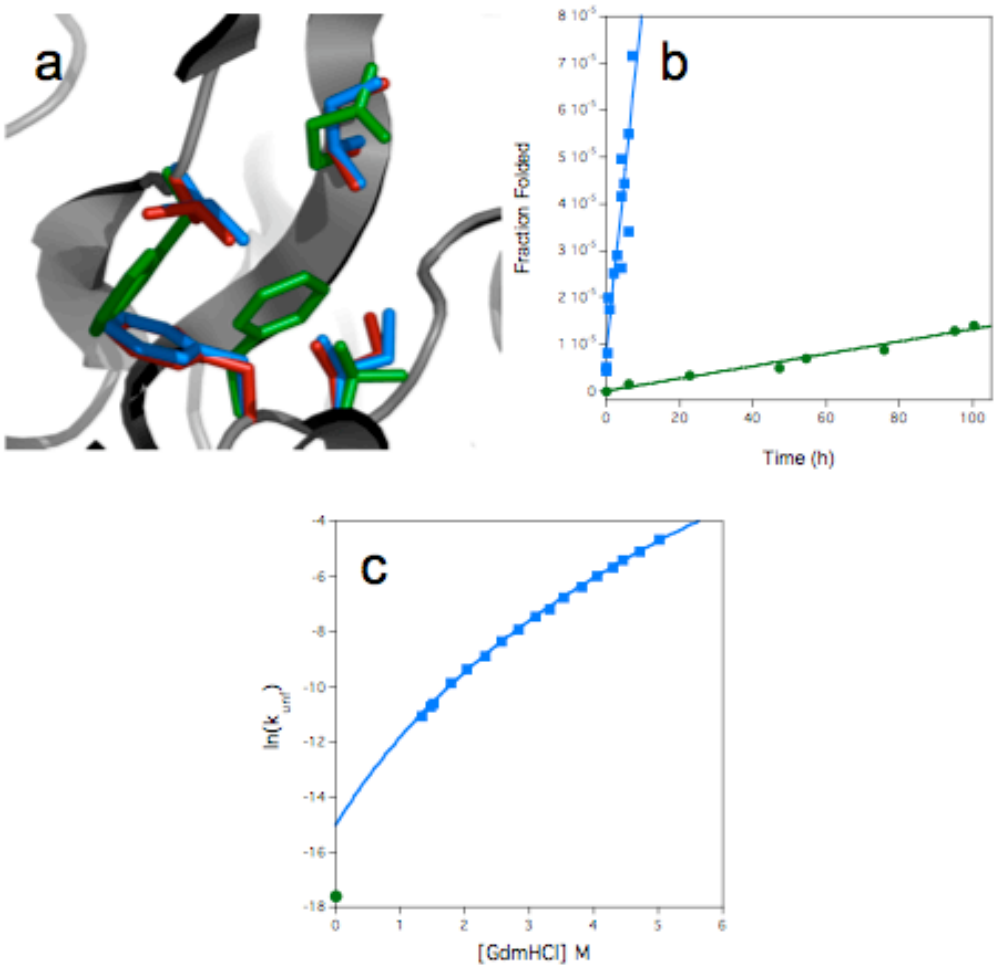


Figure 3.3: Structure and folding kinetics of Repack mutant.
A) C-domain packing of wild-type αLP(Fuhrmann, Kelch et al. 2004) (green), Repack (1.5Å;blue), and SGPB (1.7Å; red). C-domain residues of Repack superpose with those of SGPB. B) Time course of folding for WT-αLP (green), Repack (blue). Repack folds ~60 times faster than WT. C) Extrapolated unfolding rate for Repack (blue) compared to the unfolding rate for WT (green). Repack unfolds ~13-times faster than WT-αLP.

While the mutants described above suggest that the strain from sidechain distortion plays a key role in destabilizing the αLP TS, the additional removal of stabilizing Van der Waal interactions, entropic effects of mutation to glycine and the mild decrease in strain in T181A (Fig. 2A), masks the true effect of strain removal. To solve this problem, we sought to rebuild the αLP hydrophobic core so that strain is eliminated

while Van der Waals interactions are maximally preserved. Using SGPB as a guide(Huang, Lu et al. 1995), αLP was mutated (T181I, W199L, Q210I; "Repack") in order to select for the unstrained F228 rotamer. A 1.5Å structure of Repack shows that the residues in the C-domain core superpose with those of SGPB (Figure 3A) and confirms that Phe228 is in an undistorted conformation.

In accordance with our hypothesis, the Repack mutant has a huge enhancement in the folding rate as compared to wild-type (~60-fold faster; Figure 3B). while the unfolding rate is ~13-fold faster than WT (Figure 3C). The energetic consequences of an unstrained C-domain is both the TS and N are stabilized by ~2.2 and 0.8 kcal/mol, respectively (Fig. S3), provided further evidence that that strain from the Phe228 distortion is present in both N and TS. Additionally, the greater stabilization of the TS by distortion removal in Repack also suggests that, in wild-type αLP, strain is more prevalent in TS than in N. Although Repack does not fit all the underlying assumptions inherent in Phi Value analysis(Matouschek, Kellis et al. 1989), the Phi-value calculated for this mutant is a remarkable 2.7 (±0.5), clearly illustrating that the interactions in the αLP C-domain significantly contribute to the protein's enhanced kinetic stability.

The Repack folding landscape reveals that, like SGPB(Truhlar 2004), both the TS and N states are stabilized relative to WT-αLP (Fig. S3). Therefore, Repack captures key energetic, as well as structural, features of SGPB, accounting for ~36% of the SGPB TS stabilization with only ~2.5% of the sequence variation between the two proteases. Thus, the strain in the C-domain core can also partially explain the differences in the folding landscapes of SGPB and αLP.

The results presented here provide strong evidence that strain energy from the distortion of Phe228 is present both in N and TS. Because the strain destabilizes the TS to a greater extent than N, the functional role of the strain resulting in sidechain distortion is that the unfolding barrier is extended, thereby slowing unfolding (Fig. 4). Therefore, strain provides an evolutionary advantage by actually extending the functional lifetime of the protease. Because longevity increases exponentially with the unfolding barrier, kinetic stability allows relatively small increases in barrier height to profoundly affect functional lifetime. Based on the results of the Repack mutant, the strain in the C-domain core could account for a >10-fold extension of αLP lifetime. In this sense, one could think of the strain as a form of tensegrity(Fuller 1975), in which structural integrity is based on a synergy between balanced tension and compression components.

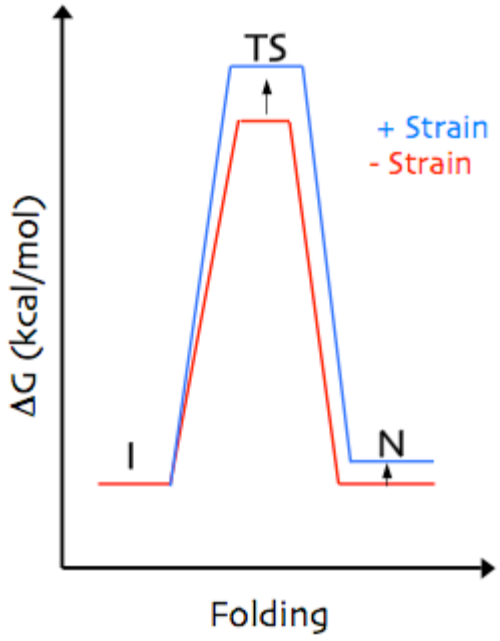
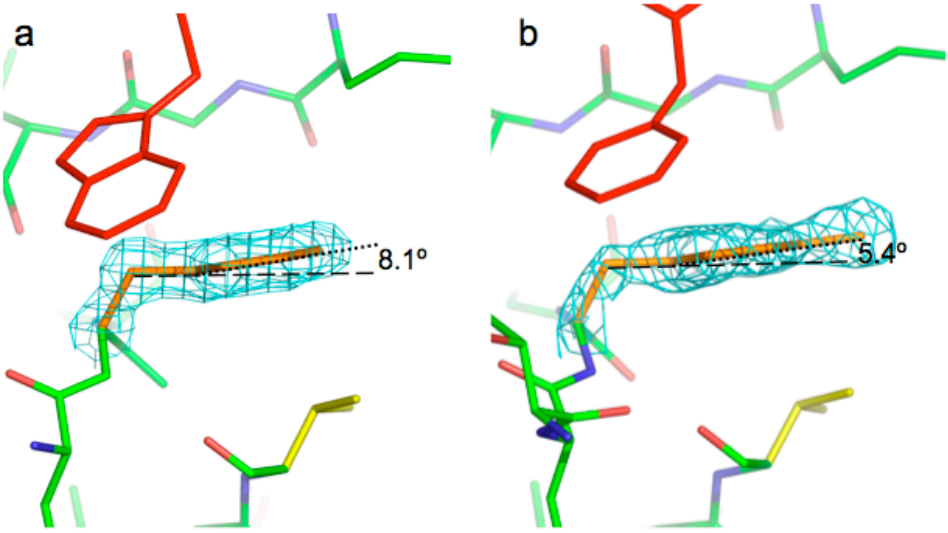


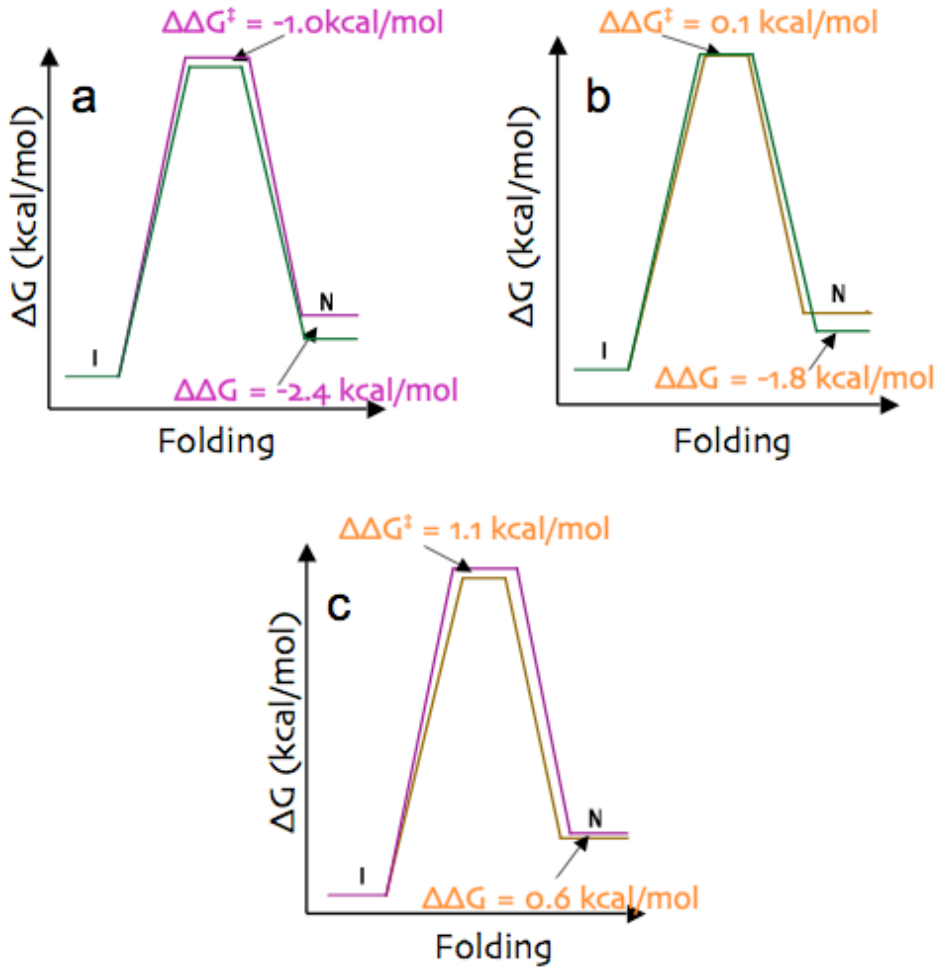
Figure 3.4: Strain increases kinetic stability.
A schematic model for the effect of strain on the folding landscape of the kinetically stable proteases. Strain destabilizes the Transition State (TS) more so than the native state (N), thereby enlarging the unfolding free energy barrier and extending protease lifetime.

To our knowledge, this is the first time that sidechain distortion has been linked to functional properties of biological macromolecules. While distortions in aromatic sidechains are relatively rare, they are not unique to αLP and its close homologs. Out of the ultrahigh-resolution structures (≤0.99Å) in the Protein Data Bank, ~20% of all proteins in our dataset of high resolution structures display a Phe distortion similar to or greater than that found in αLP (Figure 1A). The results presented here illuminate a functional advantage of sidechain distortion in αLP, suggesting that other proteins may also utilize the strain energy from sidechain distortions to modulate biological activities, such as allostery and catalysis, in addition to stability. While strain if porphyrin rings and in catalytic substrates have been shown to be functionally relevant, sidechain distortion could play a significant role in modulating energetic landscapes to provide biologically important advantages. Therefore, this study identifies an unanticipated challenge to observe structurally subtle yet functionally significant covalent distortions to fully understand the forces acting on proteins.

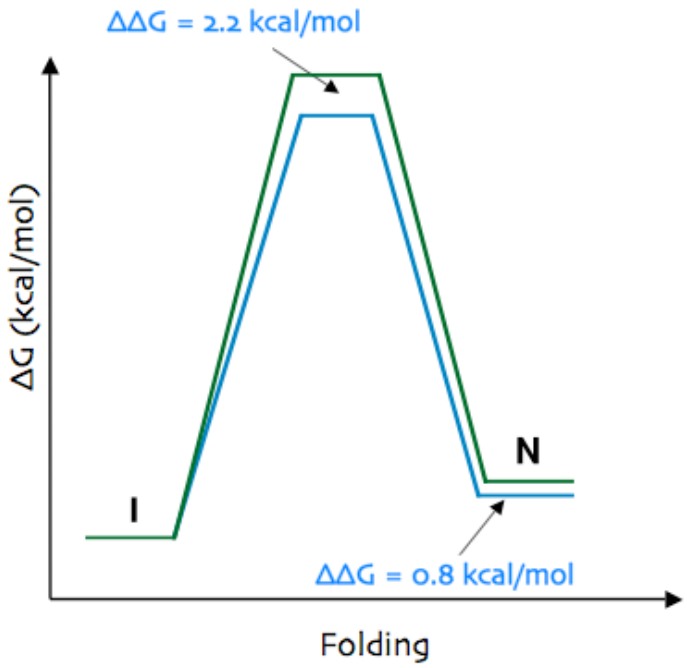
Supplemental Figures



Supplemental Figure 3.1: Distortion of Residue 228 is a conserved feature of Proteases dependent on a large Pro Region.
A) *Thermobifida fusca* Protease A (TFPA) at 1.44Å displays a large distortion in Phe228 (8.1°). B) *Nocardiopsis alba* Protease (NAPase) at 1.85Å displays a distortion in Tyr228 (5.4°).



Supplemental Figure 3.2: Free Energy Diagrams for folding of T181A (violet) and T181G (orange) compared to WT (green).
A) The T181A TS is destabilized by 1 kcal/mol and N by 2.4 kcal/mol, whereas B) the T181G TS is stabilized by 0.1kcal/mol while the N is destabilized 1.8 kcal/mol. C) Relative to T181A, T181G has a stabilized TS (1.1 kcal/mol) and N (0.6 kcal/mol).



Supplemental Figure 3.3: Folding Free Energy Diagrams for Repack (blue) compared to WT (green).
The Repack TS is stabilized by 2.2 kcal/mol and N by 0.8 kcal/mol. This is similar to SGPB, which lower barriers to folding and unfolding, and a more stabilized N than αLP.

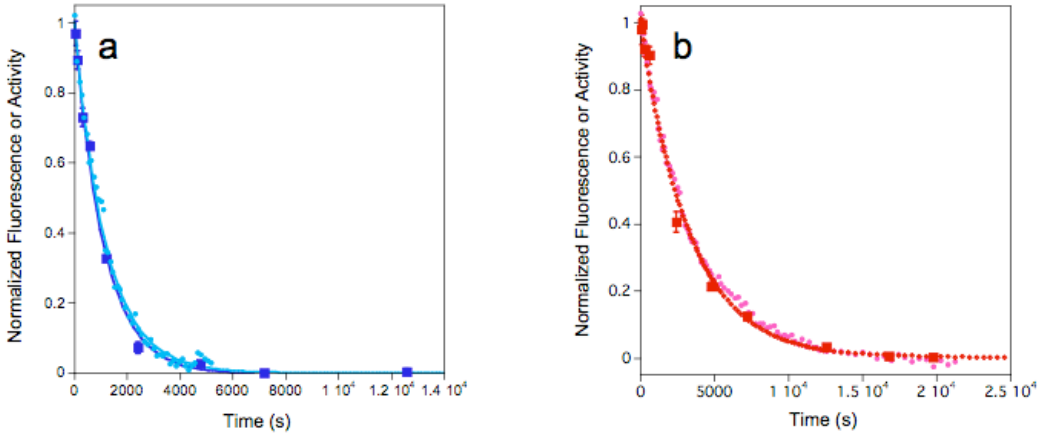
Supplemental Text

Folding of the F228A mutant

Folding of the F228A mutant was so slow as to prevent determination of an accurate rate constant ($\lesssim 5 \times 10^{-12}$ s^{-1}), which is not surprising given αLP's extremely slow folding kinetics and the drastic nature of this mutation. The large loss of stability due to the sidechain deletion efficiently masks any stabilizing effect from strain removal. Therefore, the effect of strain removal in F228A is inconclusive.

Strain and Unfolding Cooperativity

The height of the unfolding barrier is not sole determinant of lifetime in kinetically stable proteins; unfolding cooperativity is also a key feature. Sub-global structural fluctuations that could render the protein susceptible to proteolysis are completely suppressed in αLP and other homologs(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). We tested whether strain from Phe228 is acts through modulating the unfolding free energy barrier height as well as cooperativity through the use of autolysis assays. If strain removal also results in enhanced sub-global fluctuations, then autolysis (proteolysis due to other αLP molecules) will increase the rate of inactivation. In this case, αLP acts as both the enzyme and substrate, producing a highly non-linear response to protein concentration. We measured the rate of protease inactivation by loss of activity and/or fluorescence at 0.05μM and 5μM αLP at pH 8.0, close to αLP's maximal protease activity, thus increasing the rate of autolysis ~10^4-fold.



Supplemental Figure 3.4: Strain Removal has negligible effects on unfolding cooperativity.
A) Repack unfolding in 3M GdmHCl at pH8 and 0ºC. The inactivation rate of 5μM Repack (dark blue squares) is identical to that at 0.05μM Repack (light blue). B) F228A unfolding in 3M GdmHCl at pH8 and 0 ºC. The inactivation rate at 5μM F228A (dark red squares, protease activity; dark red diamonds, fluorescence) is identical to that at 0.05 μM F228A (pink circles, fluorescence).

As seen in Supplementary Figure 4, Repack and F228A show no significant difference in inactivation rates across a 100-fold difference in concentration at 3M GdmHCl and 0 ºC. This result indicates that there is no significant increase in the native state dynamics due to removal of strain in Phe228. In agreement, SGPB is degraded at its global unfolding rate(Truhlar, Cunningham et al. 2004) despite lacking distortion of Phe228(Huang, Lu et al. 1995). Therefore, the effect of strain in extending lifetime in αLP is exclusively through the height of the unfolding free energy barrier.

Materials and Methods

Database of Phe distortions

A database of atomic resolution structures (<1.0 Å) was built by examining the Brookhaven Protein Data Bank. Structures with >90% sequence identity to another entry were removed. Phenylalanine residues with multiple conformations were not included. Structures with multiple protein copies per asymmetric unit were averaged. For each Phe residue, the C_β-C_γ-C_ζ angle was measured.

Sequence analysis

Protease sequences were obtained by using BLAST(Altschul, Gish et al. 1990) of the NCBI sequence database (http://www.ncbi.nlm.nih.gov/BLAST/) with the αLP protease sequence as the query. The signal sequence was determined with the SignalP server(Bendtsen, Nielsen et al. 2004), and the Pro sequence was estimated based off of homology with the closest known relative.

MSAs were created using ClustalW(Chenna, Sugawara et al. 2003) with the BLOSUM substitution matrix(Henikoff and Henikoff 1992). Phylogenetic trees were made using the Neighbor Joining mehtod(Saitou and Nei 1987) in MacVector.

Covariation was estimated using complete MSAs of the Pro-dependent proteases. Covariation in this case is defined as differential conservation at each position dependent on the size of the corresponding pro region. In other words, the position could be conserved as one type of amino acid in the small pro proteases, but unconserved in the large pro proteases, or vice versa. Alternatively, the residue could be conserved within each class, but not conserved across the classes. Each position was visually inspected for covariation and scored accordingly.

Cloning/protein production

αLP mutants were made using Quik-Change Site-Directed mutagenesis (Stratagene) and protein was expressed and purified according to published protocols(Mace, Wilk et al. 1995).

Folding/Unfolding

Uncatalyzed folding of αLP was performed as described in Sohl *et al*(Sohl, Jaswal et al. 1998), except that the total protein concentration was ~4μM and the assay was calibrated within each timepoint using standards of no protein and 25pM wild-type αLP to improve the precision of each measurement. Unfolding was measured by loss of tryptophan fluorescence (excitation 283nm, emission 320nm) as described(Jaswal, Truhlar et al. 2005). Unfolding rate of αLP at 0 ºC was determined by extrapolation from

data of Jaswal *et al* (Jaswal, Truhlar et al. 2005). The denaturant binding model was used to model the unfolding behavior of the Repack and T181A mutants because it provides an accurate empirical fit of the curved data. Its use has been validated for wt-αLP, where it was established that the observed curvature was not due to transition state movement(Jaswal, Truhlar et al. 2005). The curvature is due to electrostatic effects from the ionic denaturant guanidine, as linear extrapolations from the non-ionic denaturant urea yield the same rate constant(Jaswal 2000). It is unknown why certain mutants have curvature in their unfolding denaturant titration, but it has been seen in multiple cases. Rate constants were converted to ΔG using standard transition state theory. Although the appropriate 'pre-exponential factor' for an unfolding or folding reaction is controversial, $\Delta\Delta G^{\ddagger}$ is unaffected by the choice of a pre-exponential term.

Crystallization/structural analysis

The crystallization of F228A, T181A and T181G αLP mutants was performed as described for wt-αLP(Fuhrmann, Kelch et al. 2004). Crystals were in space group $P3_221$ with one molecule per asymmetric unit. However, the Repack mutant could only be crystallized under slightly different conditions (addition of 10mM $CuCl_2$) and in a different space group $P6_122$ with two αLP molecules per asymmetric unit. Data was collected at the Advanced Light Source, Beamline 8.2.2 and processed in Denzo and Scalepack(Otwinowski 1997) or HKL2000(Otwinowski 1997). For the Repack mutant, the structure was solved using Molecular Replacement with Wt-αLP (with residues 181, 199, 210, and 228 as alanine) as a starting model. The F228A, T181A and T181G mutants were solved with a starting model of wild-type αLP with residues 228 and 181 as

alanine or glycine, depending on the mutant. Initial refinement was performed in CNS(Brunger, Adams et al. 1998) and anisotropic B-factors were modeled using REFMAC.(Murshudov, Vagin et al. 1997) Structural alignments were made using Combinatorial Extension(Shindyalov and Bourne 1998).

Distortion of Residue 228 in TFPA and NAPase

The distortion of Phe228 in TFPA has been described previously (BAK and DAA, submitted). The 1.85Å structure of NAPase was reported previously (Kelch et al, in press), in which distortion of Tyr228 in NAPase could be detected even with restraints (~4.2º) present. For this work, planarity restraints were removed in CNS(Brunger, Adams et al. 1998) and the distortion in the two copies of NAPase in the asymmetric unit were averaged together. To determine the degree of planarity in SGPB, we used model corresponding to 4SGB(Huang, Lu et al. 1995).

Autolysis assays

Autolysis assays were performed 0º C and pH8.0 (10mM Tris). For both Repack and F228A, unfolding was measured at 50nM protein using intrinsic tryptophan fluorescence (283nm, 322nm). Inactivation at 5μM was measured by loss of proteolytic activity toward the synthetic substrate Succinate-Ala-Pro-Ala-paranitroanilide as previously described(Jaswal, Sohl et al. 2002) and by tryptophan fluorescence (F228A only).

Chapter 4: Structural and Mechanistic Exploration of Acid Resistance: Kinetic Stability Facilitates Evolution of Extremophilic Behavior

Preface

Kinetic stability allows for extended protein lifetime, even in harsh, proteolytic environments. To examine the structural basis for this behavior, we studied the unfolding behavior of an acid-resistant αLP homolog, NAPase. I identified NAPase from a BLAST search, and pursued a collaboration with Shinji Mitsuiki to obtain NAPase protein for crystallization and biochemical characterization.

This Chapter was submitted and accepted as an article in the Journal of Molecular Biology. Dr. Shinji Mitsuiki purified the NAPase protein. Kyle Eagen, Pinar Erciyas, Elisabeth Humphris, and Adam Thomason performed the Relocation mutant unfolding and gladiator assays while under my supervision. I am responsible for all other data, all analysis, and appear as first author on this work.

Abstract

Kinetically stable proteins are unique in that their stability is determined solely by kinetic barriers rather than by thermodynamic equilibria. To better understand how kinetic stability promotes protein survival under extreme environmental conditions, we have analyzed the unfolding behavior and determined the structure of *Nocardiopsis alba* Protease A (NAPase), an acid-resistant, kinetically stable protease, and compared these results with a neutrophilic homolog, alpha-lytic protease (αLP). Although NAPase and

αLP have the same number of acid-titratable residues, kinetic studies revealed that the height of the unfolding free energy barrier for NAPase is less sensitive to acid than that of αLP, thereby accounting for NAPase's improved tolerance of low pH. A comparison of the αLP and NAPase structures identified multiple salt bridges in the domain interface of αLP that were relocated to outer regions of NAPase, suggesting a novel mechanism of acid stability in which acid-sensitive electrostatic interactions are rearranged to similarly affect the energetics of both the native state and the unfolding transition state. An acid-stable variant of αLP in which a single inter-domain salt-bridge is replaced with a corresponding intra-domain NAPase salt bridge shows a dramatic >15-fold increase in acid resistance, providing further evidence for this hypothesis. These observations also lead to a general model of the unfolding transition state structure for αLP protease family members in which the two domains separate from each other while remaining relatively intact themselves. These results illustrate the remarkable utility of kinetic stability as an evolutionary tool for developing longevity over a broad range of harsh conditions.

Introduction

Proteins from extremophilic organisms have evolved to remain stable and functional under extraordinarily harsh conditions, including extremely hot, acidic, or proteolytic environments.(Jaenicke 1991; Scandurra, Consalvi et al. 2000) Elucidating the mechanisms by which these proteins attain such remarkable stability is important for understanding both protein folding(Kumar and Nussinov 2001) and optimal engineering strategies.(Russell and Taylor 1995) Most studies are complicated by the complex, global reaction of protein structure to extreme conditions, such as high

temperature.(Jaenicke 1991; Fersht 1999) However, because acid affects protein structure at discrete distinct and specific loci, a mechanistic understanding of acidophilic adaptation can be achieved by focusing on the local environment of acid-titratable moieties in the structure of an acidophile.

Acidophilic proteins must have mechanisms in place to counter the buildup of large net positive charge due to neutralization of carboxylates,(Goto and Nishikiori 1991) the loss of carboxylate mediated salt bridges,(Chen, You et al. 1991; Ionescu and Eftink 1997) and, at moderately acidic pH, the protonation of buried histidines.(Flanagan, Garcia-Moreno et al. 1983; Hughson and Baldwin 1989; Barrick and Baldwin 1993; Geierstanger, Jamin et al. 1998) Many acidophilic proteins, such as pepsin,(Cooper, Khan et al. 1990) xylanase from *A. kawachii*,(Fushinobu, Ito et al. 1998) the soxF protein from *S. acidocaldarius,*(Bonisch, Schmidt et al. 2002) and the acid-tolerant killer toxin from *P. farinose*,(Kashiwagi, Kunishima et al. 1997) have actually evolved an overabundance of acidic surface residues, thus reducing their overall pI and their net positive charge at any pH. While this strategy increases stability at low pH by reducing positive electrostatic repulsion, there is a large increase in negative electrostatic repulsion and corresponding reduction in stability at neutral and alkaline pH.(Bender 1947; Buzzell 1952; Edelhoch 1957; Favilla, Parisoli et al. 1997) Additional strategies include reducing the overall number of charged residues as well as the solvent accessibility of acidic residues (as in acidophilic maltose binding protein from *A. acidocaldarius*(Schafer, Magnusson et al. 2004) and N-carboxyaminoimidazole ribonucleotide mutase from *A. aceti*(Settembre, Chittuluru et al. 2004)) and the binding of metal or cofactors (as in

rusticyaninin from *T. ferroxidans*(Botuyan, Toy-Palmer et al. 1996; Walter, Ealick et al. 1996)).

All of these mechanisms have been identified in proteins whose functional stability is presumably dictated by their thermodynamic stability (difference in free energy between the native and unfolded states), ΔG_{N-U}. However, for a growing list of proteins, including capsid protein SHP(Forrer, Chang et al. 2004), lipase,(Rodriguez-Larrea, Minning et al. 2006) pyrrolidone carboxyl peptidase,(Kaushik, Ogasahara et al. 2002) subtilisin BPN',(Eder, Rheinnecker et al. 1993) α-lytic protease,(Baker, Sohl et al. 1992; Sohl, Jaswal et al. 1998) and its homolog *Streptomyces griseus* Protease B (SGPB)(Truhlar, Cunningham et al. 2004), it has been shown that persistence of the native state is dictated not by ΔG_{N-U}, but by the height of their unfolding free energy barrier (ΔG_{N-TS}) which acts to kinetically trap the native state. The functionally relevant states for kinetically stable proteins are the native and transition states, both of which can be well structured. Conversely, in thermodynamically stable proteins, the differential stability between the structured native state and the unstructured unfolded state determines global protein stability. We hypothesize that these different stability mechanisms provide unique opportunities that kinetically stable proteins can exploit when developing resistance to extreme conditions.

To understand the interplay between kinetic stability and acid resistance, we focused on the well-understood, kinetically stable bacterial proteases of the chymotrypsin superfamily. These proteases display extraordinarily slow folding ($t_{1/2,folding}$ = 1800 years(Sohl, Jaswal et al. 1998) for αLP and 3 days for SGPB(Truhlar, Cunningham et al. 2004)) and unfolding rates ($t_{1/2,unfolding}$ = 1.2 years for αLP(Sohl, Jaswal et al. 1998) and 11

days for SGPB(Truhlar, Cunningham et al. 2004)). Remarkably, the αLP native state is actually less stable than unfolded forms by an unprecedented ~4 kcal/mol.(Sohl, Jaswal et al. 1998)

To facilitate folding on a biological timescale, these proteases have co-evolved a covalently attached pro region that catalyzes the folding of the protease.(Peters, Shiau et al. 1998; Sauter, Mau et al. 1998; Truhlar, Cunningham et al. 2004) After folding has completed, the pro region is proteolytically destroyed, yielding the kinetically-trapped native protease.(Cunningham and Agard 2004) Importantly, the transient involvement of the pro region uncouples the folding and unfolding landscapes, which allows evolution of unique and beneficial native state properties.(Jaswal, Truhlar et al. 2005) For example, the αLP and SGPB native states are extremely rigid, which limits both global and local unfolding processes, thereby significantly reducing their susceptibility to proteolytic degradation and increasing functional lifetime.(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004)

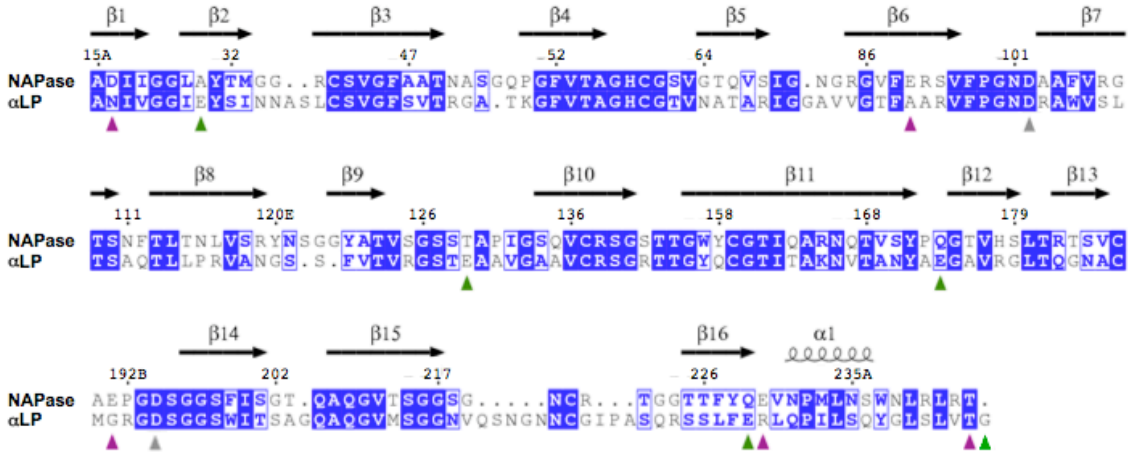


Figure 4.1: Aligned sequences of NAPase and αLP with secondary structure diagramed above.

Identities are shown as blue boxes with white lettering and similarities as blue lettering in hollow blue boxes. Sequence alignment is based on aligned structures of αLP and NAPase.(Shindyalov and Bourne 1998) Violet triangles denote acidic residues unique to NAPase, while green triangles are those unique to αLP. Residues are numbered according to homology to chymotrypsin.(Fujinaga, Delbaere et al. 1985)

Nocardiopsis alba Protease A (NAPase), from the bacterium *N. alba strain TOA-1* which was isolated from a bathroom tile joint(Mitsuiki, Sakai et al. 2002), was identified as a secreted, keratinolytic protease with considerable stability under acidic conditions.(Mitsuiki, Sakai et al. 2002) The strong sequence homology with αLP (44% identical, 64% homologous; see Figure 4.1) and the presence of a very similar pro region provide key hallmarks to identify NAPase as a kinetically stable protease. In the studies presented here, we find that NAPase unfolds more slowly than αLP at high temperature and that the unfolding of NAPase is less sensitive to acidic conditions, resulting in a significantly enhanced functional lifetime at low pH. However, unlike other acid stable enzymes, NAPase has the same total number of acidic residues (six) as its acid labile counterpart, αLP. A structural and mutational comparison of NAPase and its acid-labile homolog, αLP, simultaneously provides strong evidence for a novel mechanism of acid resistance through charge migration and leads to a new model of the unfolding transition state structure for these proteases.

Results

Unfolding at High Temperature

As a first step toward characterizing the NAPase energetic landscape, we measured its unfolding rate at 70 ºC and pH 5. It is possible to avoid the confounding pH and chemical effects typical of using other denaturants by using high temperature to

accelerate the unfolding reaction. The rate of tertiary structure loss as measured by protease activity (1.31 (±0.05) x 10^{-4} s^{-1}; Figure 4.2a) closely parallels the rate of secondary structure loss as measured by circular dichroism (1.61 (±0.05) x 10^{-4} s^{-1}; Figure 4.2b), consistent with a two-state unfolding reaction . The rate of αLP unfolding under identical conditions, as measured by loss of protease activity and fluorescence, is 2.47 (±0.01) x 10^{-3} s^{-1}. This is 17 times faster than NAPase (Figure 4.2a), showing that the unfolding of NAPase is less sensitive to high temperature than αLP.

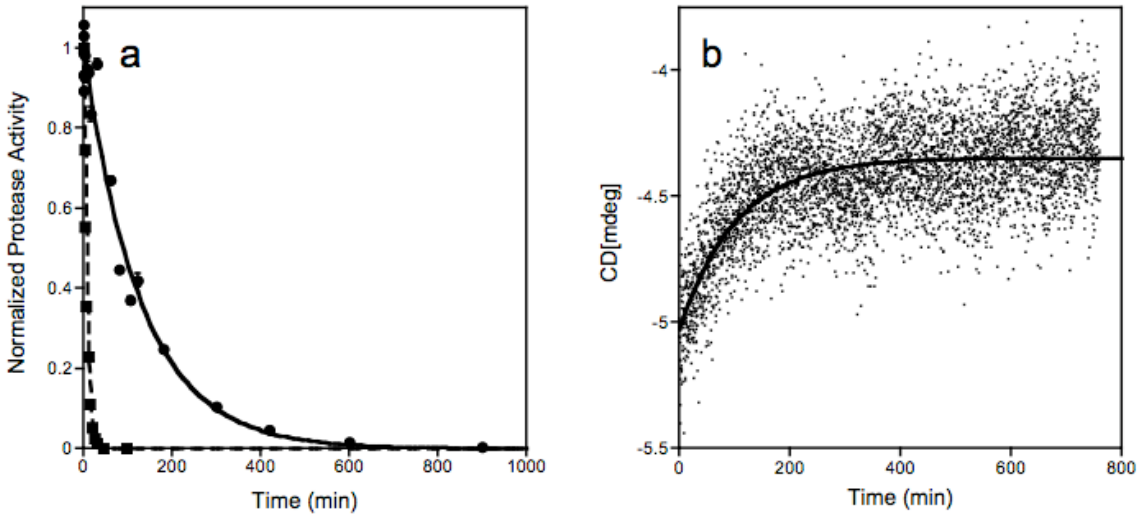


Figure 4.2: Unfolding of NAPase (circles, solid line) at 70 °C and pH 5.0 followed (a) by loss of protease activity and (b) by circular dichroism.
The data are fit using a single exponential equation yielding similar rate constants of 1.31 (±0.05) x 10^{-4} s^{-1} and 1.61 (±0.05) x 10^{-4} s^{-1}, respectively, suggesting unfolding by a two-state mechanism. As shown in (a), αLP (squares, dashed line) unfolds under identical conditions with a rate constant of 2.46 (±0.01) x 10^{-3} s^{-1}.

NAPase is more acid-resistant than αLP

Since NAPase has previously been shown to be qualitatively resistant to acid,(Mitsuiki, Sakai et al. 2002) we quantified the effect of varying pH on the unfolding

rates of both NAPase and αLP. To accurately determine unfolding rate constants, we chose temperatures (αLP 60 ºC, NAPase 70 ºC) where the unfolding rate constants were similar for αLP ($t_{1/2}$ ~ 4.7h) and NAPase ($t_{1/2}$ ~ 1.3h).

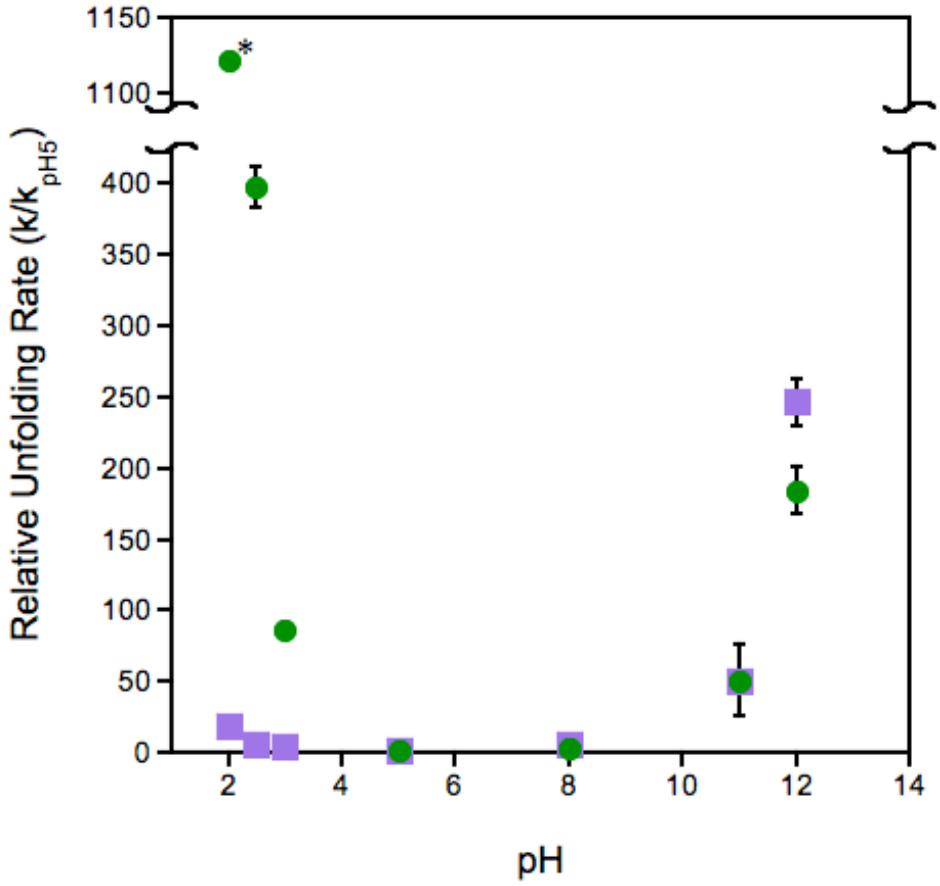


Figure 4.3: Effect of pH on unfolding rate for αLP (green) and NAPase (violet). The y-axis represents the increase in unfolding rate relative to pH 5.0. Unfolding was followed by circular dichroism and/or loss of protease activity. Unfolding was triggered by high temperature: 60 °C for αLP and 70 °C for NAPase.

The pH dependence of unfolding differed substantially between αLP and NAPase (Figure 4.3 and Supplemental Table 4.1). The unfolding rate of NAPase is relatively insensitive to low pH; at pH 2.0, NAPase unfolds only ~20-fold faster than at pH 5.0,

whereas at pH 12 it is accelerated ~175-fold. In contrast, the rate of αLP unfolding is extremely sensitive to low pH (~400-fold faster at pH 2.5 than at pH 5.0). In fact, unfolding of αLP at pH 2.0 was so fast ($t_{1/2}$ < 15s) as to prevent determination of an accurate unfolding rate constant using manual mixing to initiate unfolding. Thus at pH 2.5, αLP is ~80-fold more acid sensitive than NAPase (relative acid sensitivity = $(k_{pH2.5, \alpha LP}/k_{pH5, \alpha LP})/ (k_{pH2.5, NAPase}/k_{pH5, NAPase})$). Altogether, NAPase is extraordinarily resistant to extremes of both heat and acid; the $t_{1/2}$ for unfolding at pH 2.5 and 70 ºC is a remarkable 15.5 minutes, while for αLP it is estimated to be ~0.7 seconds, a ~1300-fold difference in unfolding rate under identical conditions.

To obtain insight into the nature of the unfolding cooperativity of NAPase as a function of pH, we analyzed the resistance to proteolysis of αLP and NAPase. No matter how protease resistant the native state, local as well as global unfolding can lead to proteolytic accessibility, and hence inactivation. In principle, unfolding cooperativity could be so extreme as to completely suppress local fluctuations so that only global unfolding events lead to proteolytic destruction. Indeed, proteolysis has been shown to be a useful tool for examining sub-global fluctuations of proteins in a manner analogous to hydrogen exchange experiments.(Rupley 1967; Park and Marqusee 2004; Park and Marqusee 2005) Proteolytic susceptibility is conveniently determined using a 'survival assay',(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004) in which multiple proteases are mixed together in equimolar ratios at 37 °C and proteolysis is allowed to proceed. Because αLP, NAPase and trypsin have orthogonal substrate specificities (data not shown), one can readily measure the amount of active protease remaining over time

by measuring the activities of each enzyme using three different chromogenic substrates, each of which is specific for a different protease.

The results for the survival assay at pH 7.0 and 37 ºC (Figure 4.4a) show that αLP outlasts NAPase by a factor of ~2, a significant difference considering the impressive longevity of both of these pro-dependent proteases. The loss of trypsin activity is faster by factors of about 20 and 10 than the rates of degradation for αLP and NAPase, respectively. In contrast, the survival assay at pH 2.5 shows an opposite trend for αLP and NAPase (Figure 4.4b). While the degradation rates at pH 2.5 for both proteases are increased relative to those at neutrality, the $t_{1/2}$ of αLP inactivation is 1.4±0.2 days while that for NAPase is 12±6 days, or roughly a 9-fold difference in inactivation rates. This result illustrates that the superior ability of NAPase to resist acidic conditions extends over a broad range of temperatures and in highly proteolytic environments.

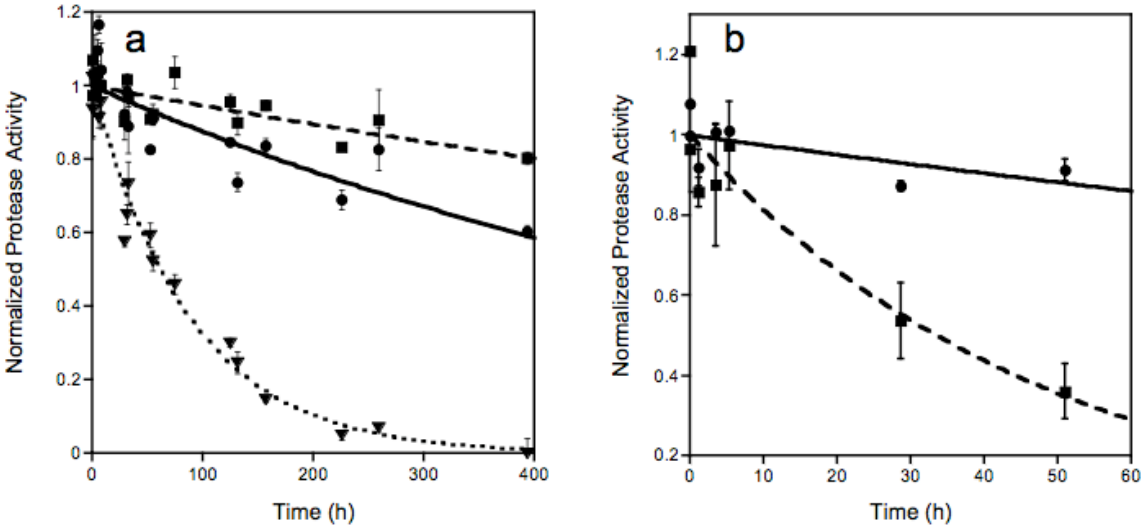


Figure 4.4: pH dependence of protease survival.
Co-incubation of αLP (squares) and NAPase (circles) at 37 ºC with (a) trypsin (triangles) at pH 7.0 or (b) pepsin (pepsin data not shown) at pH 2.5. (a) At pH 7.0, NAPase is degraded about two-fold faster than αLP but about 10-fold slower than trypsin ($k_{deg, \alpha LP, pH7}$

= 1.5 (± 0.2) x 10^{-7} s^{-1}, $k_{deg, NAPase, pH7}$ = 3.7 (± 0.6) x 10^{-7} s^{-1}, $k_{deg, tryspin, pH7}$ = 3.1 (± 0.4) x 10^{-6} s^{-1}). (b) However, αLP is degraded faster than NAPase at pH 2.5 ($k_{deg, \alpha LP, pH2.5}$ = 5.7 (± 1.2) x 10^{-6} s^{-1}, $k_{deg, NAPase, pH2.5}$ = 6.9 (± 2.8) x 10^{-7} s^{-1}). All data were fit with a single exponential equation.

The Structure of NAPase

To gain insight into the mechanisms generating acid stability in NAPase, we determined the structure of NAPase by x-ray crystallography to a resolution of 1.85Å (Table 4.1). There were two identical copies of NAPase in the crystallographic asymmetric unit (C_α RMSD = 0.20Å), and all atoms of the protein were visible in electron density maps with the exception of two arginine sidechains past the C_β atoms (Arg220 and 241 [residues numbered according to chymotrypsin homology using scheme of Fujinaga, *etal*.(Fujinaga, Delbaere et al. 1985)]). NAPase exhibits the double β-barrel architecture characteristic of chymotrypsin family members. Like other Pro region dependent proteases of this family,(James, Sielecki et al. 1980; Nienaber, Breddam et al. 1993; Kitadokoro, Tsuzuki et al. 1994; Huang, Lu et al. 1995; Fuhrmann, Kelch et al. 2004) NAPase is distinguished from the classical chymotrypsin fold by two structural elements: i) a small β-hairpin (termed the "domain bridge") that connects the N- and C-terminal β-barrels (residues 120A to 120L), and ii) a larger β-hairpin in the C-terminal domain (residues 156 to 185), shown to be an integral site for Pro region binding and folding catalysis.(Peters, Shiau et al. 1998; Sauter, Mau et al. 1998; Truhlar and Agard 2005) As expected from the high level of sequence conservation, the overall structure of NAPase is extremely similar to that of αLP(Fuhrmann, Kelch et al. 2004) (C_α RMSD =

1.2Å and 1.1Å for NAPase molecules A and B, respectively; see structural overlay in Supplemental Figure 4.1).

Table 4.1: NAPase Structure Statistics

Data Statistics	
Space group	P 3_2 2 1
Unit cell dimensions (Å)	$a = b = 65.9$, $c = 79.7$
Molecules per asu	2
Limiting resolution (Å)	1.85
I/σ(I)	16.7 (3.9)[a]
Completeness (%)	97.4 (98.7)[a]
R_{merge}[b](%)	9.7 (46.9)[a]

Structure Refinement	
Resolution range (Å)	50 – 1.85
R (%)	17.6
R_{free} (%)	20.0
# of Protein residues[c]	376
# of Waters[c]	391
# of AEBSF ligands[c]	2
# of Sulfate ions[c]	2
# of Glycerol[c]	2
# of Dioxane[c]	2
Average isotropic B-factors	
Protein atoms (Å^2)[c]	12.1 (±4.9)
Solvent atoms (Å^2)[c]	36.0 (±16.4)

[a]Shown in parentheses for the highest resolution bin, 1.88 to 1.85Å
[b]R_{merge} as calculated by Scalepack(Otwinowski 1997)
[c] Values given for the asymmetric unit, which contains two NAPase molecules

Despite the strong structural homology, there are regions in which the backbone traces of NAPase and αLP significantly deviate (Supplemental Figure 4.1). The largest backbone change occurs at a loop in the C-terminal domain (residues 216 to 225). The

extended loop found in αLP is unique in the sub-family and has been shown to play a role in dictating substrate specificity.(Mace and Agard 1995; Mace, Wilk et al. 1995) Nearly all other Pro region dependent family members, including NAPase, have a much smaller and more conserved loop. Given its high degree of conservation it is unlikely that this loop would be the key to NAPase's unique acid resistance. Another region of NAPase with large backbone changes relative to αLP is the 'domain bridge', a β-hairpin that connects the N- and C-terminal domains (residues 120A to 120L). At both a structural and sequence level, the domain bridge of NAPase most closely resembles that of a thermophilic homologue, *Thermobifida fusca* Protease A (TFPA) (BAK & DAA, submitted). Preliminary studies suggest that the domain bridge and its packing interactions may contribute to the greater thermostability of TFPA and NAPase (BAK & DAA, submitted).

Electrostatic Features of the NAPase Structure

The structure of NAPase provides a unique opportunity to gain insight into the mechanisms of acid-resistance through comparing its electrostatic attributes with those of the homologous, yet acid-labile, αLP. NAPase and αLP have the exact same number of acidic residues (six) and a comparable percentage of positive residues (NAPase: 6.91%, αLP: 7.58%), which is inconsistent with the two previously identified general mechanisms for acid stability.(Kashiwagi, Kunishima et al. 1997; Fushinobu, Ito et al. 1998; Bonisch, Schmidt et al. 2002; Schafer, Magnusson et al. 2004; Settembre, Chittuluru et al. 2004) Similarly, metal binding site effects proposed to stabilize the *T. ferrooxidans* rusticyanin in acid(Botuyan, Toy-Palmer et al. 1996; Walter, Ealick et al.

1996) are irrelevant for NAPase, which contains no metals or other cofactors (as evidenced by its structure), suggesting that NAPase utilizes a unique strategy for acid stability.

In the relevant pH range (<5.0), αLP and NAPase each contain seven acid-titratable moieties (6 acidic residues and the C-terminal carboxylate) and form a similar number of salt bridges (7 in NAPase, 8 in αLP; Table 2). Two of these residues (Asp102 and Asp194) are critically important for catalytic function(Nakagawa and Umeyama 1984; Craik, Roczniak et al. 1987; Sprang, Standing et al. 1987; Carter and Wells 1988; Hedstrom, Lin et al. 1996; Reiling, Krucinski et al. 2003) and are conserved in nearly all chymotrypsin family serine proteases, and thus are unlikely to be relevant for acid stability. The remaining five carboxylates occupy different positions in both the primary sequence (Figure 4.1) and tertiary structure (Figure 4.5), and are thus good candidates for the differential low pH stability of αLP and NAPase.

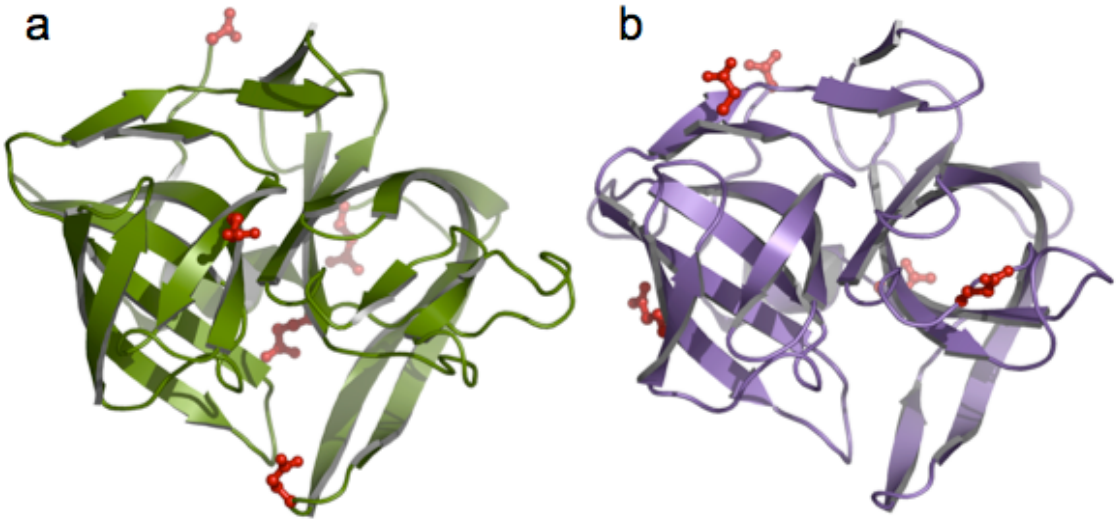


Figure 4.5: Distribution of acidic residues in (a) αLP (green) and (b) NAPase (violet).

Non-conserved acidic residues are shown in red (ball and stick models). Compared to αLP, the acidic residues of NAPase have migrated away from the domain interface.

In αLP, all carboxylates are located at or near the interface between the N- and C-domains, whereas in NAPase only the C-terminus is located near this interface (Figure 4.5). Despite strong overall sequence conservation for residues in the region of the domain interface, the distribution of charges, both negative and positive, across this interface is strikingly different between αLP and NAPase (Figure 4.6). In particular, all of the inter-domain salt bridges in αLP (37.5% of the total salt bridges) are replaced with hydrophobic or polar interactions in NAPase, which has no inter-domain ion pairs at all (Figure 4.6). Additionally, the NAPase ion pairs involve residues that are close in primary sequence (average sequence separation = 20 residues), whereas those in αLP are significantly more distant (average sequence separation = 78 residues). In fact, NAPase has three salt bridges involving adjacent residues (Asp15B-Nterm, Glu88A-Arg89, and Cterm-Arg243) while αLP contains none (Table 4.2).

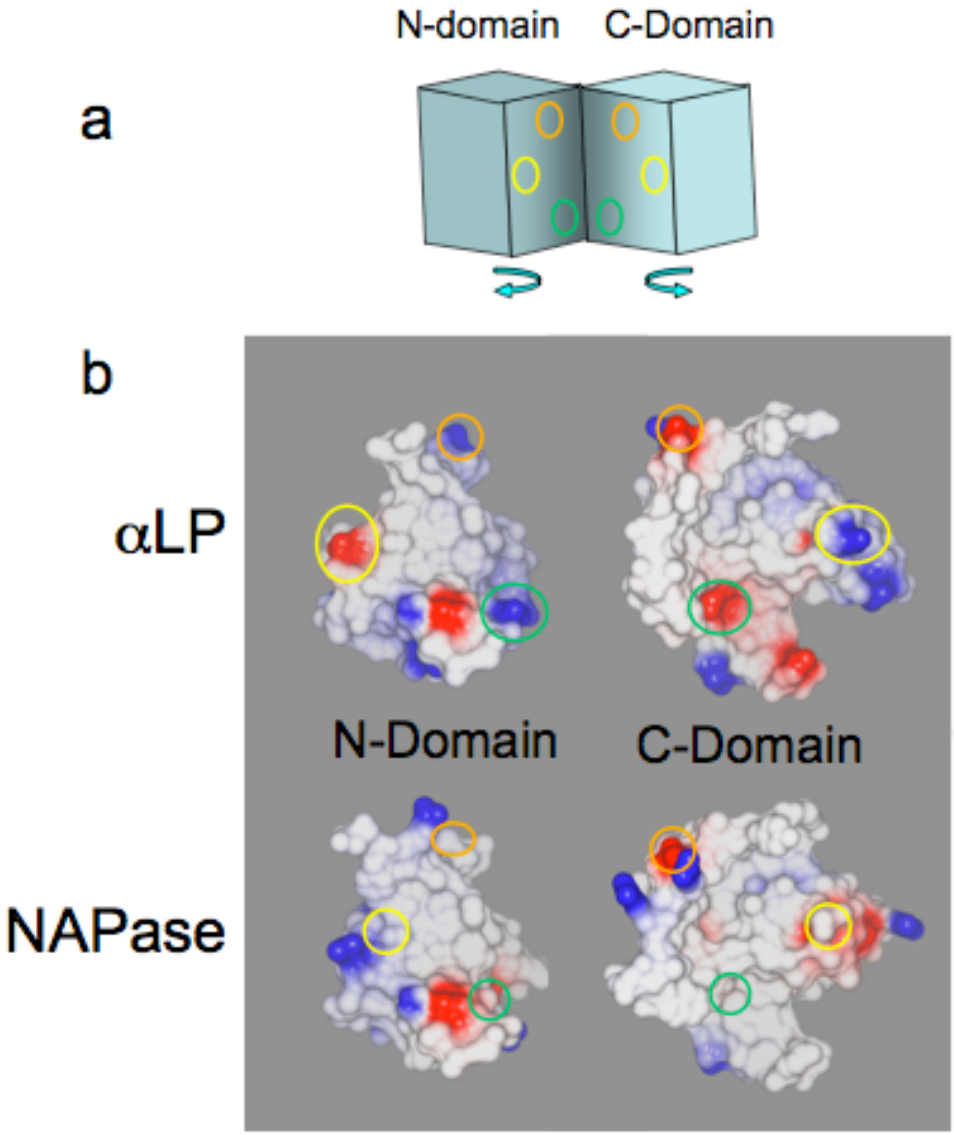


Figure 4.6: Electrostatic properties of the domain interfaces of αLP and NAPase. (a) Illustrates the orientation of the molecular surfaces shown in (b), cracked open at the domain interface. (b) Formal charge is displayed on the surface of the buried domain interfaces for αLP (top) and NAPase (bottom), revealing areas of positive (blue) and negative (red) electrostatic potential. Circles highlight residues in inter-domain salt-bridges of αLP that are absent in NAPase. Each salt bridge pair is denoted by a different circle color. Electrostatic potentials were calculated and displayed using ccp4MG.(Potterton, McNicholas et al. 2002; Potterton, McNicholas et al. 2004)

Table 4.2: A Comparison of Non-conserved Salt Bridges

αLP Salt Bridge	Inter-Domain?	Sequence Separation	NAPase Cognate Residues[a]	Type of replacement Interaction in NAPase[b]
Glu32…Arg141	Yes	97	Ala, Ser	VdW, H-bond
Glu129…Lys165	No	24	Thr, Arg	Salt Bridge, H-bond
Glu129…Arg230	No	90	Thr, Glu	Salt Bridge, H-bond
Glu229…Arg103	Yes	118	Gln, Ala	VdW, H-bond
Cterm…Arg122	No	109	Cterm[c], Thr	Salt Bridge
Cterm…Arg120A	Yes	120	Cterm[c], Leu	Salt Bridge, VdW

NAPase Salt Bridge	Inter-Domain?	Sequence Separation	αLP Cognate Residues[a]	Type of replacement Interaction in αLP[b]
Asp15B…Nterm	No	1	Asn, Nterm	H-bond
Glu88A…Arg89	No	1	Ala, Ala	VdW
Glu88A…Arg107	No	13	Ala, Ser	VdW
Glu230… Arg165	No	57	Arg, Lys	Salt Bridge, H-bond
Cterm…Arg243	No	1	Cterm[c], Val	Salt Bridges, VdW

[a]Amino acids that replace residues involved in the salt bridge, respectively
[b]Interactions that the cognate residues make with each other or with other parts of the protein
[c]The NAPase C-terminal residue (Thr244) is penultimate of the C-terminus of αLP (Gly245).

Salt Bridge Relocation Mutant

To further investigate the role of this alternate distribution of salt bridges in acid stability, we mutated three residues in αLP to relocate electrostatic interactions within the protease. We converted an inter-domain salt bridge (Glu32 with Arg141) to those residues found in NAPase (Glu32➔Ala32, Arg141➔Ser141), therefore abolishing this salt bridge and simultaneously recreating a NAPase intra-domain ion pair (Nterm with

Asp15B) within αLP (Asn15B➔Asp) to conserve the total number of ion pairs (Relocation Mutant; Figure 4.7a). The location of the replacement salt bridge was chosen so as to be far from other αLP ionic groups to avoid introducing spurious electrostatic interactions. The Relocation Mutant unfolds ~6-fold faster than WT-αLP at pH 5.0 and 60 ºC (Supp. Table 4.1), indicating that the overall unfolding free energy barrier has been lowered. However, when the acid sensitivity of the unfolding rate was examined, the Relocation Mutant was found to be much less sensitive to acid, having only a ~60-fold acceleration at pH 2.0 relative to pH 5.0 (Figure 4.7b). Altogether, the Relocation Mutant is ≥20-fold more acid stable than WT-αLP at pH 2.0 and ~15-fold more stable at pH 2.5 (increase in acid stability = $(k_{pH5,\ mutant}/k_{low\ pH,\ mutant})/(k_{pH5,\ WT}/k_{low\ pH,\ WT})$; Figure 4.7b).

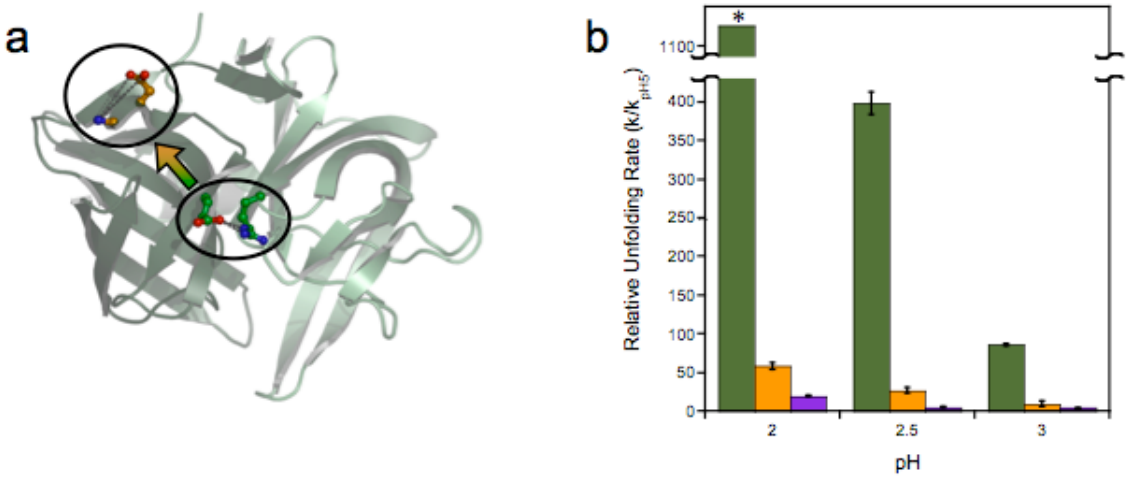


Figure 4.7: Relocation of salt bridges increases acid stability of αLP.
(A) Illustrates the changes made in αLP to generate the Relocation mutant. Glu32 and Arg141 are mutated to Ala and Ser, respectively, abolishing this inter-domain salt bridge (shown in green). Asn15B is mutated to Asp to create a new, intra-domain ion pair with the N-terminus (orange), as found in NAPase. (B) Effect of low pH on the unfolding rate of the Relocation mutant. The Relocation mutant (orange) is >15fold more acid stable than αLP (green), and approaches that of NAPase (violet).

Discussion

Acid Resistance via Kinetic Stability

For both αLP and NAPase, the rate of protein unfolding is the sole determinant of functional 'stability'. While NAPase does exhibit enhanced high temperature resistance as evidenced by a slower unfolding rate at 70 °C (Figure 4.2a), here we focus on understanding the much larger difference (~80-fold) in the sensitivity of the unfolding rates to acid (Figure 4.3). As shown, the reduced responsiveness of the unfolding free energy barrier to low pH directly translates into enhanced longevity for NAPase under very acidic conditions (Figure 4.4b). Because the NAPase unfolding rate does not increase significantly with acidification (Figure 4.3), low pH must either (1) not perturb the energetics of the native (N) and transition states (TS) or (2) alter the energetic of both states to a similar extent, such that the relative energy difference remains essentially unchanged (Figure 4.8a). Although we do not know what the effect of acid is on the stability of N with respect to U, it does not matter for NAPase longevity and function. The unfolding free energy barrier still retains the same height, thus, NAPase is kinetically acid stable (Figure 4.8a). In contrast, the pronounced acid sensitivity observed for αLP indicates that exposure to acid has a significant differential effect on N and TS (Figure 4.8b), resulting in a reduced unfolding free energy barrier.

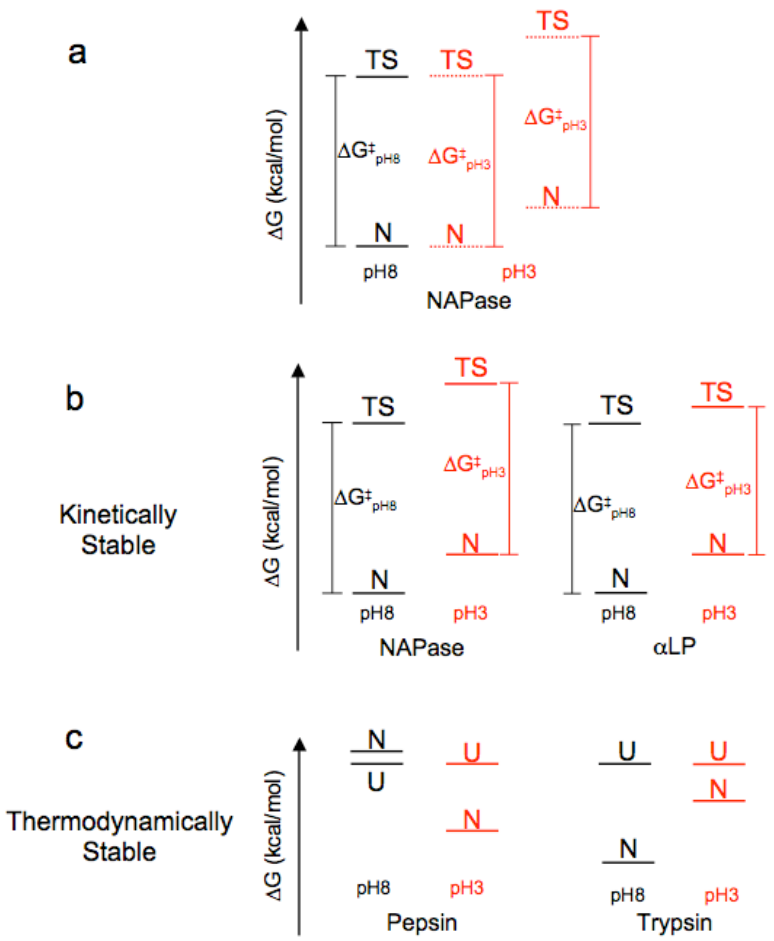


Figure 4.8: Qualitative schematic illustrating the effect of acid on the NAPase unfolding free energy barrier (A) and different strategies for acidophilic behavior in (B) kinetically and (C) thermodynamically stable proteins.
Free energies are displayed at both neutral (black) and at low (red) pH. (A) NAPase is kinetically acid-stable. Because the unfolding free energy barrier maintains the same height at neutral and low pH, acid must either not perturb the energetics of the NAPase N and TS (first set of dashed lines) or alter the energetics to a similar extent (second set of dashed lines). However, acid is expected to destabilize the native state, so we have displayed it as such in the following diagram. (B and C) NAPase is contrasted with αLP (B), and pepsin with trypsin (C). (B) In kinetically stable proteins, there are differential effects on the TS of an acidophile (NAPase) versus a neutrophile (αLP). The TS of αLP is not destabilized by acid as greatly as N, which decreases the unfolding free energy barrier, whereas the TS of NAPase is destabilized to roughly the same degree as N, thereby maintaining the unfolding free energy barrier height across a wide pH range. (C) At neutral pH, trypsin is quite stable, whereas pepsin is unstable. The reverse is true at acidic pH.

This mechanism stands in marked contrast to the acid-stabilization strategies used by thermodynamically stable proteins which include minimizing electrostatic repulsion (by increasing the number of carboxylates and thereby lowering the pI),(Cooper, Khan et al. 1990; Kashiwagi, Kunishima et al. 1997; Fushinobu, Ito et al. 1998; Bonisch, Schmidt et al. 2002) or reducing the number of carboxylates and thus reducing the magnitude of the pH dependent effect.(Schafer, Magnusson et al. 2004; Settembre, Chittuluru et al. 2004) However, a major consequence is that protein stability is often seriously compromised at neutral pH.(Bender 1947; Buzzell 1952; Edelhoch 1957; Favilla, Parisoli et al. 1997) This is illustrated in the schematic energy landscape diagrams in Figure 4.8c, where trypsin (stable at neutral pH) and pepsin (stable at low pH) are contrasted at both low and neutral pH.

Given the previous data on acid stability mechanisms, it was surprising to see that NAPase is not exceptionally sensitive to neutral or basic conditions (Figure 4.3) and that both NAPase and αLP have the same number of acid titratable groups and a very similar number of salt bridges at neutral pH. Increased salt bridge content also seems to be correlated with thermostability(Karshikoff and Ladenstein 2001), yet this correlation also does not hold for αLP and NAPase. Taken together, these behaviors show that NAPase must employ a very different acid stability mechanism and indicate that this is a likely consequence of its kinetic stability. This raises several key questions: (1) What is the mechanism of NAPase's extraordinary acid stability? (2) What can this information tell us about the structural origins of kinetic stability? and (3) How can it remain stable over such a broad pH range?

Acid Stability through Salt Bridge Migration

Our structural analysis indicates that, despite the proteins' overall conservation, the locations of salt bridges in αLP and NAPase are very different. Foremost, in NAPase there is a migration of acid sensitive residues (Figure 4.5), and therefore salt bridges (Figure 4.6), away from the interface of the N- and C-domains. This result suggests that the domain interface plays a key role in the N to TS transition (see below). Furthermore, the salt bridges in NAPase (all intra-domain) occur between residues that are much more local in terms of the primary sequence of the protein (i.e. low contact order(Plaxco, Simons et al. 1998)) than in αLP (Table 4.2). In fact, NAPase has three salt bridges involving adjacent residues whereas αLP has no such interactions. These highly localized ion pairs are more likely to remain interacting in the TS due to their close sequence proximity. Therefore, their electrostatic interaction energy will tend to be closely matched between the N and TS, so that acid-induced salt bridge destabilization would have minimal effects on the height of the unfolding free energy barrier (Figure 4.8b). Because these ion pairs are likely interacting in both N and TS, they actually would provide no net kinetic stabilization, but may be present to increase protein solubility. Conversely, the acid sensitivity of αLP suggests that its salt bridge arrangement must be significantly disrupted between N and TS.

A Model for the Transition State of Kinetically-Stable Proteases

The presence of inter-domain, acid-labile salt bridges in αLP, but not NAPase, would suggest that the interactions along this interface are broken during the N-to-TS

transition. Consequently, acid-induced disruption of αLP's inter-domain salt bridges would destabilize N but would have a lesser effect on the energetics of the TS. In NAPase, by contrast, relocation of acid-labile interactions away from the interface would make the N-to-TS transition relatively insensitive to acid, since the N and TS states would be affected equally.

To further test the role of inter-domain salt bridges in αLP, we used NAPase as a guide for relocating one of the three αLP inter-domain salt bridges to the periphery (Figure 4.7a). The mutation was surprisingly potent, accounting for more than 60% of the differential acid stability of αLP and NAPase despite representing less than 3% of their sequence variation. This data strongly supports the proposal that intra-domain salt bridges must be broken in the αLP TS.

These results provide critical information on the global TS structure for αLP. Previously, it was proposed that there must be a relatively large change in exposed surface area but relatively little change in structure(Jaswal 2000) to explain the favorable enthalpy and unfavorable entropy of the N to TS transition.(Jaswal, Truhlar et al. 2005) Further, an analysis of the change in solvent exposed hydrophobic surface area during unfolding showed that the TS is very native-like (~80% native).(Jaswal 2000) These results dictate a TS structure in which there is significant ordering of solvent without a large loss of protein structure. Combining this with our new salt bridge data suggests an αLP TS model in which both β-barrel domains separate from each other but remain relatively intact themselves, allowing water molecules to become ordered in the newly made crevice between the two domains (Figure 4.9).

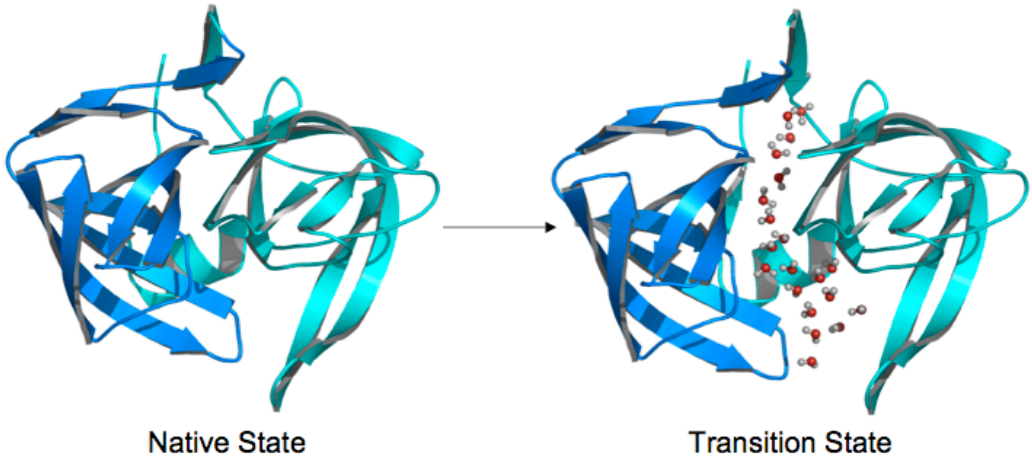


Figure 4.9: Proposed Unfolding TS Model for Pro region dependent proteases.
The N- and C-domains themselves remain relatively intact but separate from each other, allowing solvent to enter into the inter-domain crevice.

As further corroboration, preliminary results from molecular dynamics simulations of αLP unfolding suggest that interactions along the domain interface are among those predominantly broken during the N to TS transition [N. Salimi, DAA, unpublished results.] Finally, the proposed TS structure could also help explain αLP's overall resistance to proteolysis,(Jaswal, Sohl et al. 2002) as the protein only enters a protease-susceptible conformation (i.e. exposure of loops) after it has globally unfolded. This relatively well-folded TS may have been evolved to protect the protease from proteolysis, illustrating how the details of a folding landscape can have important implications for protein function.

While the majority of our data provides structural insights into the αLP TS, several factors argue that the model should be broadly applicable to other kinetically stable family members. First, the structural data presented here demonstrates that the native states are very similar. Second, it has been demonstrated that the folding pathways

of homologous proteins within several protein families are quite similar.(Stackhouse, Onuffer et al. 1987; Kragelund, Hojrup et al. 1996; Perl, Welker et al. 1998; Chiti, Taddei et al. 1999; Clarke, Cota et al. 1999; Martinez and Serrano 1999; Schindler, Graumann et al. 1999; Gunasekaran, Eyles et al. 2001; Hollien and Marqusee 2002; Vallee-Belisle, Turcotte et al. 2004) Third, where examined, the folding transition states for kinetically stable αLP family members have been found to be quite compact and very native-like.(Jaswal, Sohl et al. 2002; Truhlar 2004) Fourth, the αLP folding TS seems to be a highly specific and very resilient structure. Mutations effect the catalyzed (i.e. in the presence of the pro region) and uncatalyzed folding reactions equivalently even though their rates differ by $> 10^9$.(Derman and Agard 2000) Together these data provide strong support for a TS structural model (Figure 4.9) that pertains not only to αLP but to NAPase and other kinetically stable family members.

Kinetic Stability Facilitates Protein Adaptation

While most acidophilic proteins have evolved to reside solely in low pH environments (and, therefore, don't need to survive and/or function under neutral or alkaline conditions; Figure 4.8b), NAPase is most active at neutral or high pH and normally exists in an environment (tile grout) that is quite basic.(Mitsuiki, Sakai et al. 2002) However, as with many biofilm forming microorganisms, *Nocardiopsis alba* and other *Nocardiopsis* species can greatly acidify their environment in response to certain nutrients(Lejbkowicz, Kudinsky et al. 2005). Thus the creation of a local, and perhaps transient, low pH environment around the colony could necessitate the development of

secreted proteins like NAPase that are capable of retaining activity and stability at both high and low pH.

Unlike most other adaptations involving alterations in number of acid titratable groups, the combination of kinetic stability and salt bridge migration provides stability over a large pH range. Since the unfolded state (U) is inconsequential for stability, a kinetically stable protein can allow an arbitrary destabilization of the native state with no consequences for longevity, provided that the TS is equally destabilized (Figure 4.8a). This is easily realizable as only a small number of interactions differ between the N and TS in these proteins (compared to N and U for thermodynamically stable proteins). Therefore, kinetically stable proteins may employ mechanisms for adaptation to a harsh environment that are not available to thermodynamically stable proteins. While kinetic stability is shown here to stabilize NAPase over broad pH conditions, this concept should be generalizable for development of resistance to other destabilizing agents, such as alkali, heat, or chemical denaturants (eg. urea or guanidine). Kinetic stability provides the potential for rapid evolutionary change to accommodate environmental changes and therefore can be viewed as a flexible paradigm for attaining extraordinary longevity in extremely harsh environments and over broad ranges of conditions.

Materials and Methods

Protein Production

NAPase was expressed and purified as described previously.(Mitsuiki, Sakai et al. 2002) Lyophilized NAPase was resuspended in 2 mM sodium acetate, pH 5.0 and buffer exchanged to remove residual salt. NAPase for use in crystallization experiments was

treated with the serine protease inhibitor aminoethylbenzylsulfonyl fluoride (AEBSF). 100 molar excess AEBSF in 100mM HEPES, pH 7.0 was added to concentrated NAPase and this solution was incubated at room temperature overnight. Near 100% inhibition of NAPase was obtained. NAPase-AEBSF was then concentrated in a BioMax 5 kDa molecular weight cutoff filter (Millipore) and buffer exchanged with 2mM sodium acetate, pH5.0 in the concentrator. αLP was prepared as previously described.(Mace, Wilk et al. 1995) All chemicals were purchased from Sigma Aldrich.

The Relocation mutant (Asn15B->Asp, Glu32-> Ala, Arg141->Ser) was created using the QuikChange Multi Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA) with primers 5'-GCTGCAGACCACGGCCGACATCGTCGGCGGCATCGCATACTCGATCAACA-3' and 5'-GGTGTGCCGCTCGGGCAGCACCACCGGTTACCAG-3'. Mutant expression and purification were performed as described previously.(Mace, Wilk et al. 1995)

Unfolding

NAPase unfolding was measured by either circular dichroism (CD) monitored at 213nm or loss of unfolding activity. Although NAPase has two tryptophans, unfolding was not monitored by fluorescence due to low signal for unfolding under the conditions used. For CD experiments, concentrated protein was mixed with buffer pre-equilibrated to the proper temperature (final [protein] = 5μM) in a 0.2-cm pathlength cuvette and the CD signal at 213nm was monitored by a Jasco J-175 spectropolarimeter (Easton, MD). For loss of activity experiments, concentrated protein was mixed with buffer pre-equilibrated to the proper temperature (1μM final protein). Aliquots were taken at various time points and the unfolding reaction was quenched on ice. Protease activity was

subsequently measured as previously described,(Jaswal, Sohl et al. 2002) except NAPase activity was monitored using 1mM succinyl-Ala-Ala-Pro-Phe-pNA (Bachem, Bioscience Inc.) as substrate. αLP unfolding was monitored by both loss of protease activity and by fluorescence as previously described.(Jaswal, Truhlar et al. 2005) Unfolding of the Relocation mutant was similarly monitored by fluorescence with a Fluoromax-3 fluorometer (JY Horiba) (excitation 283nm, emission 322nm). The buffers (all at 10mM) utilized were: glycine (pH 2.0, 2.5 and 3.0), potassium acetate (pH 5.0), Tris (pH8.0), CAPS (pH11.0), and arginine (pH12.0). Unfolding rate constants were obtained by fitting data to a single exponential equation or with a single exponential plus a linear term.

Survival Assay

The survival assay at pH 7.0 was performed as described previously(Truhlar, Cunningham et al. 2004) except that NAPase was used in place of SGPB and final protein concentrations were 5μM. The survival assay at pH 2.5 was performed in a similar manner except that the reaction was buffered with 50mM glycine, pH 2.5 and Pepsin A (Worthington Biochemical Corp.) was utilized in place of bovine trypsin because it is much more active at low pH than serine proteases such as trypsin, NAPase and αLP. Since pepsin is unfolded and inactive at the pH of the activity assay (pH 8.0),(Bender 1947; Buzzell 1952; Edelhoch 1957; Favilla, Parisoli et al. 1997) the proteolytic activities measured for both αLP and NAPase are unaffected by the presence of pepsin. The inactivation rates for NAPase and αLP at pH 7.0 were determined from a linear fit, whereas the inactivation rates for trypsin and αLP (at pH2.5) were determined with a single exponential fit.

NAPase Crystallization and Structure Determination

Crystallization experiments were setup using the hanging-drop, vapor diffusion method in 24-well trays (Nextal Biotechnologies, Montreal, Canada) at room temperature. Thin, rod-like crystals of NAPase-AEBSF (10mg/ml) grew over a solution of 1.3M ammonium sulfate, 10% (v/v) dioxane, 0.1M MES, pH 6.0 overnight. For data collection, crystals were frozen in mother liquor with 20% (v/v) glycerol.

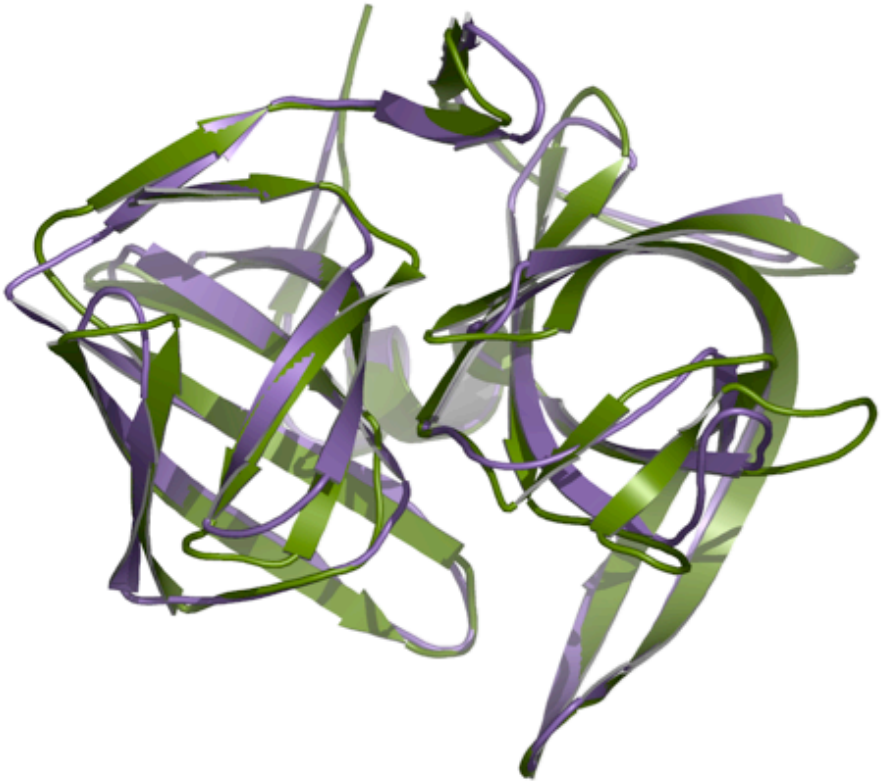
Diffraction data was collected at Advanced Light Source Beamline 8.2.2. Diffracting crystals were space group $P3_221$ with two NAPase molecules per asymmetric unit. Data reduction was performed using HKL2000(Otwinowski 1997) and initial phases were obtained using molecular replacement and a poly-Ala search model from the closely related (57% identical) protein *Thermobifida fusca* Protease A (TFPA) [BAK & DAA, manuscript in preparation.] To minimize model bias, divergent loops were deleted from the search model. Electron density maps of the protein were high quality, as all backbone atoms and all but two sidechains could be unambiguously modeled (no interpretable density beyond C_β was found in F_o-F_c difference maps for both Arg 241 or Arg 220.) Contrary to the sequence information in the NCBI database (http://www.ncbi.nlm.nih.gov/), which lists residue 196 as alanine, from the density maps this residue is unambiguously glycine. Both $2F_o$-F_c and F_o-F_c maps show no density corresponding to a C_β atom and the residue adopts Phi/Psi angles that are only allowed for glycine. Moreover, this residue, which is directly adjacent to the catalytic serine, is also conserved as glycine in nearly all known serine proteases of the chymotrypsin fold.

Refinement was accomplished using CNS(Brunger, Adams et al. 1998) and manual rebuilding using O.(Jones 1990) According to the validation program PROCHECK,(Laskowski, Macarthur et al. 1993) the resulting model had acceptable or better than average geometry, with 86.7% of residues in the most favored Ramachandran region, 12.6% in the allowed region, and 0.7% in the unfavorable region. Electron density for the two residues in the disallowed region of the Ramachandran plot (Tyr120E from both molecule A and B) was extremely clear. Structural alignments, used for Supp. Fig. 1 and the sequence alignment in Figure 4.1, were performed using the Combinatorial Extension method.(Shindyalov and Bourne 1998) Figures 4.5 and 4.9 were produced with Pymol(DeLano 2002) and Figure 4.6 with CCP4-MG.(Potterton, McNicholas et al. 2002; Potterton, McNicholas et al. 2004) The NAPase structure shown in Figure 4.6 includes two arginine sidechains which could not be accurately fit into the final model. For the purposes of this figure, these sidechains were placed into the most favorable rotamer. Since both Arg220 and Arg241 are on the periphery of the protein, the precise positioning of these residues has negligible effects on the electrostatics of the domain interface.

Acknowledgements

The authors thank Drs. C. Fuhrmann and L. Rice for assistance with many aspects of diffraction data collection, processing, and refinement. We also thank Dr. Corie Ralston at the ALS Beamline 8.2.2 for assistance in data collection. We greatly appreciate Q. Justman and Drs. L. Rice, C. Fuhrmann, and S.M.E. Truhlar for critical reading of the manuscript and members of the Agard Lab for insightful discussions.

B.A.K. and F.P.E. were supported by Howard Hughes Medical Institute Predoctoral Fellowships and K.P.E. was supported by the UCSF 2006 Summer Research Undergraduate Training Program. This work was funded by the Howard Hughes Medical Institute.



Supplemental Figure 4.1: Structural homology of αLP and NAPase.
NAPase is shown in violet and αLP in green.

Supplemental Table 4.1: Unfolding Rate Constants

pH	$k_{NAPase,\ 70°}$ (s^{-1})	$k_{\alpha LP,\ 60°}$ (s^{-1})	$k_{\alpha LP,\ 70°}$ (s^{-1})	$k_{Relo\text{-}\alpha LP,\ 60°}$ (s^{-1})
2.0	2.83 (±0.16) x 10^{-3}	ND (>0.05)	ND	0.0150 (±0.002)
2.5	7.5 (±1.4) x 10^{-4}	0.0164 (±0.0006)	ND	6.90 (±0.45) x 10^{-3}
3.0	5.49 (±0.32) x 10^{-4}	3.56 (±0.06) x 10^{-3}	ND	2.42 (±0.64) x 10^{-3}
5.0	1.46 (±0.13) x 10^{-4}	4.12 (±0.01) x 10^{-5}	2.47 (±0.01) x 10^{-3}	2.56 (±0.17) x 10^{-4}
8.0	7.63 (±0.99) x 10^{-4}	1.04 (±0.06) x 10^{-4}	ND	ND
11.0	7.31 (± 0.65) x 10^{-3}	2.09 (±1.0) x 10^{-3}	ND	ND
12.0	3.60 (±0.23) x 10^{-2}	7.58 (±0.68) x 10^{-3}	ND	ND

Chapter 5: Investigating the Electrostatic Contribution to Kinetic Stability

Preface

Analysis of NAPase has yielded a remarkable amount of insight into the benefits of kinetic stability and the structural mechanisms that create this behavior. These studies have focused on the electrostatics that govern kinetic stability in these proteases and lead to their unique behavior. To gain more detailed knowledge into the electrostatic components of kinetic stability, we used mutagenesis and kinetic studies varying ionic strength to examine the Debye-Huckel effects on unfolding.

Kyle Eagen performed all the experiments described in this chapter while working as a summer undergraduate researcher under my direct supervision. Both he and I are responsible for all analyses.

Introduction

The comparison of αLP and NAPase has highlighted the important role of electrostatics in governing kinetic stability. In NAPase, the electrostatics have been tuned through evolution to provide maximal longevity under extremely acidic conditions. A preliminary investigation of the electrostatics that underlie this adaptation provided deep insights into the structural mechanism of kinetic stability and led to structural model for the unfolding Transition State (TS; see previous chapter). Therefore, a more thorough interrogation of the electrostatic contribution to kinetic stability should provide a more

detailed understanding of the structural and mechanistic underpinnings of the extreme folding landscapes exhibited by these proteases.

Towards this end, we made mutants of αLP that recapitulated different electrostatic interactions of NAPase to determine their role in acid resistance and the unfolding barrier. These studies validated our hypothesis for the mechanistic basis of NAPase's kinetic acid stability and provided direct evidence for the structural model for the unfolding TS. In addition, we investigated the effect of salt on the unfolding transition for NAPase, αLP and some Relocation mutants. The results and implications of these studies are discussed.

Results and Discussion

To test the contribution of specific ionic interactions to the unfolding barrier to its response to acid, we created three mutants of αLP in which certain ion pairs were relocated in different positions using NAPase as a guide: Relocation1 (Asn15B to Asp, Glu32 to Ala, and Arg141 to Ser; the 'Relocation mutant' of Chapter 4), Relocation2 (Ala88A to Glu, Arg103 to Ala, Ser107 to Arg, Glu182 to Gln), and Relocation3 (Arg120A to Leu, Val243 to Arg, Gly245 to Stop codon) (Figure 5.1). The positions were chosen to avoid spurious electrostatic interactions with other endogenous αLP charge pairs. Relocation3 did not express to high enough levels for further studies, possibly because of the drastic nature of truncating αLP by one residue.

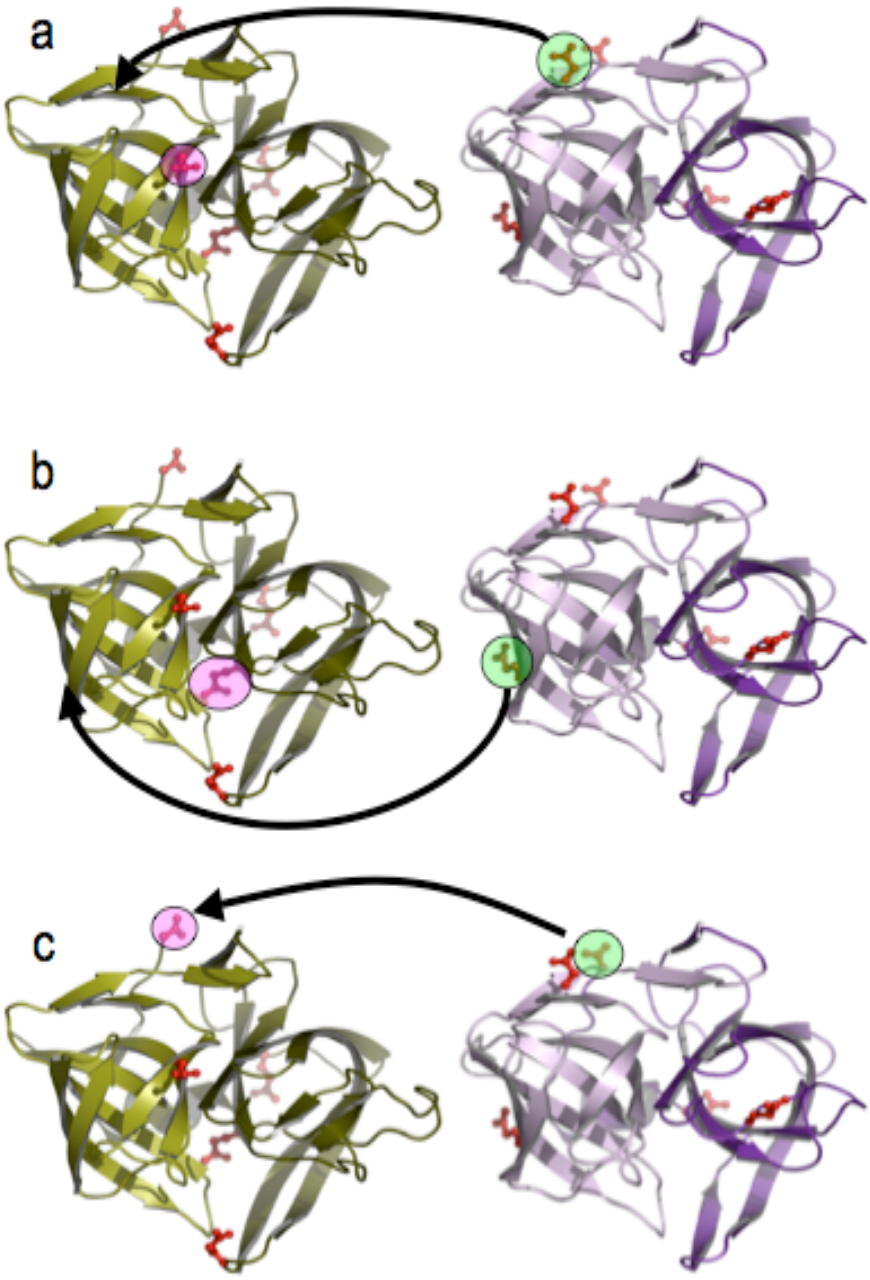


Figure 5.1: Structural diagrams of Relocation mutants.
αLP is shown in green and NAPase in violet, and all unique acidic residues are shown in red. The salt bridge associated with the violet circled residue is removed whereas the salt bridge associated with the green circled residue is created in αLP. A) The Relocation1 mutant (Asn15B→Asp, Glu32→Ala, Arg141→Ser) B) The Relocation2 mutant (Ala88A→Glu, Arg103→Ala, Ser107→Arg, Glu182→Gln) C) The Relocation3 mutant (Arg120A→Leu, Val243→Arg, Gly245→Stop codon)

It was hypothesized that the recapitulation of ion pairs from NAPase within αLP would increase kinetic acid stability. Therefore, the effect of acid on the unfolding free energy barrier of Relocation mutants 1 and 2 was investigated by measuring the unfolding rate constant using tryptophan fluorescence under various different pH values. To accelerate the reaction and avoid the confounding electrostatic effects of denaturants, high temperature (60 ºC) was used.

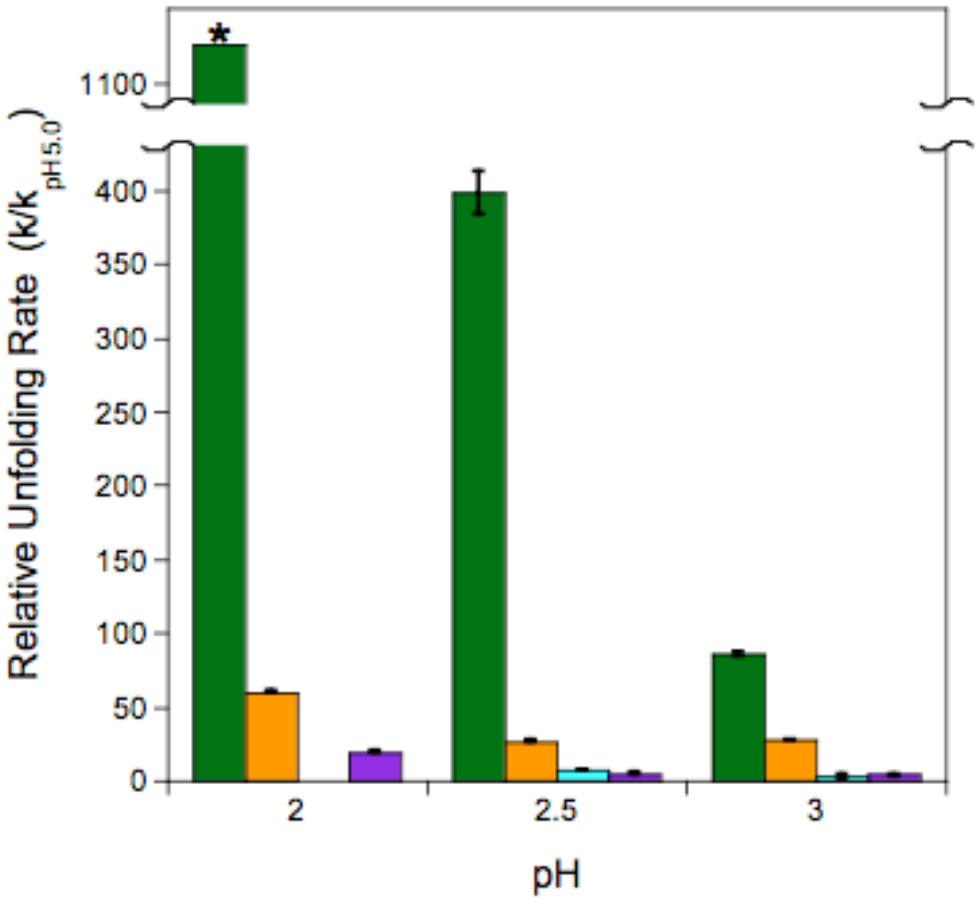


Figure 5.2: The pH dependence of unfolding for the Relocation mutants.
WT-αLP is shown in green, Relocation1 in orange, Relocation2 in cyan and NAPase in violet. The Relocation2 rate at pH2 was not measured. The rate increase for WT-αLP at pH 2.0 is a lower limit.

The Relocation1 and 2 mutants unfold ~6-fold and ~150 faster than WT-αLP at pH 5.0 and 60 ºC, indicating that the overall unfolding free energy barrier is lowered for both proteins. In addition, both Relocation mutants were found to be much less sensitive to acid, providing strong evidence that the location of these ion pairs dictate acid sensitivity. The Relocation1 mutant has only a ~25-fold acceleration in unfolding rate at pH 2.5 relative to pH 5.0, while for Relocation2 it is only ~7-fold (Figure 5.2). Altogether, Relocation1 and Relocation2 are ~15- and ~50 more stable at pH 2.5, respectively (increase in acid stability = $(k_{pH5,\ WT}/k_{low\ pH,\ WT})/(k_{pH5,\ mutant}/k_{low\ pH,\ mutant})$).

While the results for both mutants provide strong evidence that the acid sensitivity is dictated by the positioning of the ion pairs in these proteases, the magnitude of the increase in acid kinetic stability in Relocation3 is quite surprising. The relocation of a single inter-domain salt bridge causes this mutant to attain similar acid resistance as NAPase, while it was believed that the loss of each ion pair would only partially account for NAPase's acidic kinetic stability. It is not entirely clear why this mutant has such enhanced acid resistance. Mutants in which the endogenous αLP salt bridge is removed but the NAPase salt bridge not incorporated (and vice versa) could shed some light onto the individual effects of the electrostatic changes made in Relocation2.

To interrogate in further detail the electrostatic contributions to kinetic stability, we investigated the effect of salt on the unfolding free energy barrier of αLP and NAPase at pH 5. In the absence of any specific binding events, the effect of moderate salt concentrations (<1M) will be a screening of both stabilizing and destabilizing electrostatic interactions through Debye-Huckel effects(de Los Rios and Plaxco 2005).

Higher concentrations will have large effects on the solvent structure beyond electrostatic screening due to Hofmeister effects on solvent(Hofmeister 1888).

Using the non-chaotropic salt NaCl, the unfolding free energy barrier of αLP is linearly dependent on the square root of ionic strength (Figure 5.3, as predicted by the Debye-Huckel limiting law(Debye 1923). The rate of unfolding increases, indicating that salt is screening out stabilizing interactions. Conversely, the NAPase unfolding barrier is nearly independent of ionic strength, with a slight decrease in the unfolding rate with increasing ionic strength. These results agree quite nicely with our structural model of the TS for these proteins. In αLP, inter-domain salt bridges are broken in the unfolding TS. Salt acts to weaken these interactions in N, thereby lowering the unfolding barrier and accelerating unfolding. In contrast, NAPase doesn't contain interdomain ion pairs and is therefore not expected to have a large electrostatic component to its unfolding barrier, as evidenced by its lack of salt dependence.

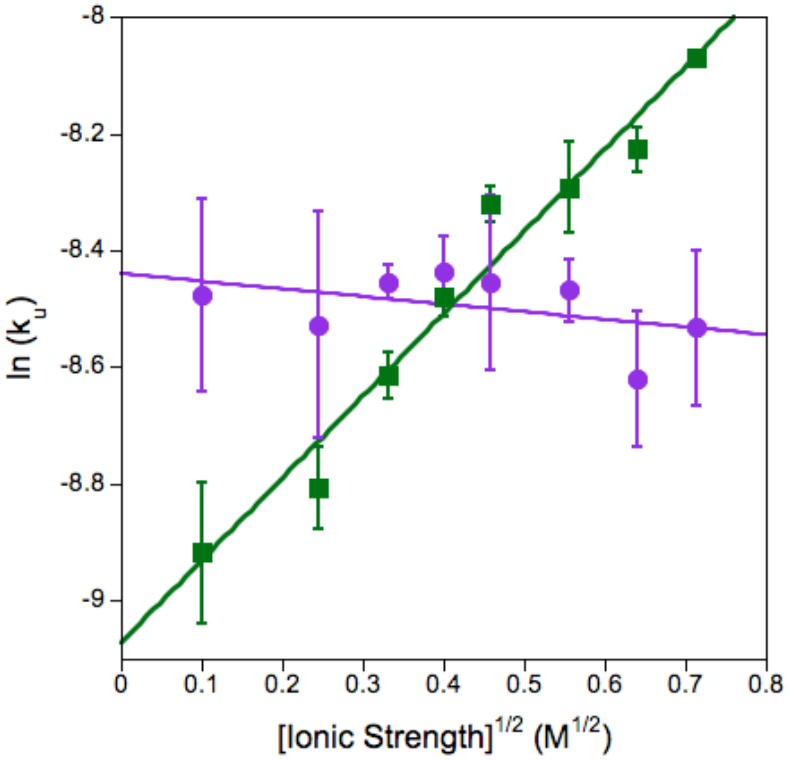


Figure 5.3: The dependence of ionic strength on pH 5 unfolding of αLP (60 ºC) and NAPase (70 ºC).
The acceleration of αLP unfolding with increasing ionic strength indicates that electrostatic screening is lowering the unfolding barrier, while the NAPase barrier is largely insensitive to electrostatic screening.

The Debye-Huckel analysis of αLP and NAPase at pH 5.0 illustrated its potential for readily addressing important issues into the electrostatic contributions to kinetic stability. We therefore sought to extend this analysis to other conditions and mutants to better understand the kinetic stability. Our hypothesis predicts that the acidic residues in the salt bridges in N that contribute to unfolding are no longer negatively charged at low pH. Therefore, we predicted that the salt dependence of unfolding would be greatly attenuated or, if salt begins to screen destabilizing positive-positive interactions, even reversed under acidic conditions.

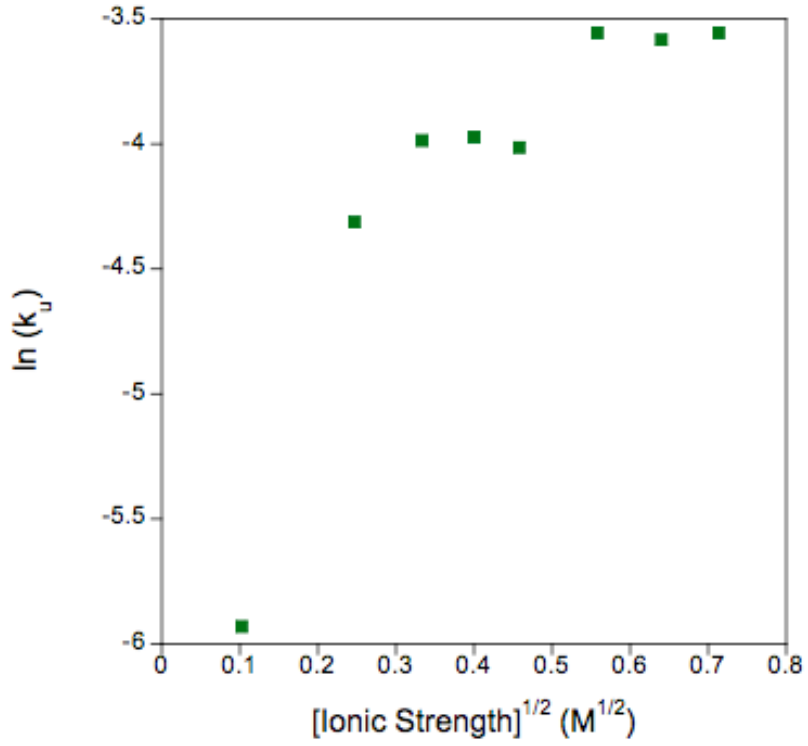


Figure 5.4: The dependence of αLP unfolding on ionic strength at pH3.
The hyperbolic dependence suggests that ion binding is playing a role in the acceleration of unfolding rate at pH3

We tested this hypothesis by measuring the unfolding rate of αLP at pH 3 with varying concentration of NaCl. Surprisingly, the unfolding rate does not exhibit an attenuated dependence on ionic strength, but increasing salt accelerates unfolding faster at low pH than at pH 5 (Figure 5.4), in direct contradiction of our hypothesis. Interestingly, the unfolding rate seems to exhibit saturable behavior with increasing NaCl concentration, which may be indicative of ion binding. This is quite remarkable and would be suggestive of preferential ion binding in the TS over N at low pH. These types of specific ion binding events have been observed previously and actually used as a sensitive probe of the structural details of the TS(Krantz and Sosnick 2001).

To investigate the specificity of ion binding in further detail, we measured the unfolding rate of αLP at pH 5 with different non-chaotropic salts: NaCl, NaBr, KCl, and KBr. If there is no specific ion binding occurring, then the unfolding rate dependence on ionic strength with the different salts should be identical. Alternatively, if specific ions are preferentially bound by N or TS, then we would expect the dependence to be attenuated or steeper, respectively.

The unfolding rates using KCl are nearly coincident with those of NaCl (Figure 5). However, the αLP unfolding rates are much higher when using NaBr or KBr (Figure 5). These results are highly suggestive that specific ion binding is occurring, and that ion binding is occurring preferentially in TS (thus stabilizing TS and lowering the unfolding barrier.) The bromide ion seems to lower the unfolding barrier because both salts that have accelerated unfolding contain this same ion. Therefore, we would hypothesize that there is a specific binding of bromide by the TS. However, this analysis does not preclude specific binding of chloride ion in the TS, but it does establish that the effect from bromide is greater than chloride. This effect is probably not due to differences in the placement of these ions on the Hofmeister series, because these differences are small and proteins in which the salt dependence of folding/unfolding was due to pure Debye-Huckel screening exhibit no difference with these salts(de Los Rios and Plaxco 2005). These salt studies suggest that the stark differences in pH 5 salt dependence exhibited by αLP and NAPase (Figure 5.3) may not be exclusively due to the screening of the αLP intedomain salt bridges, but perhaps a mixture of Debye-Huckel electrostatic screening and specific ion binding.

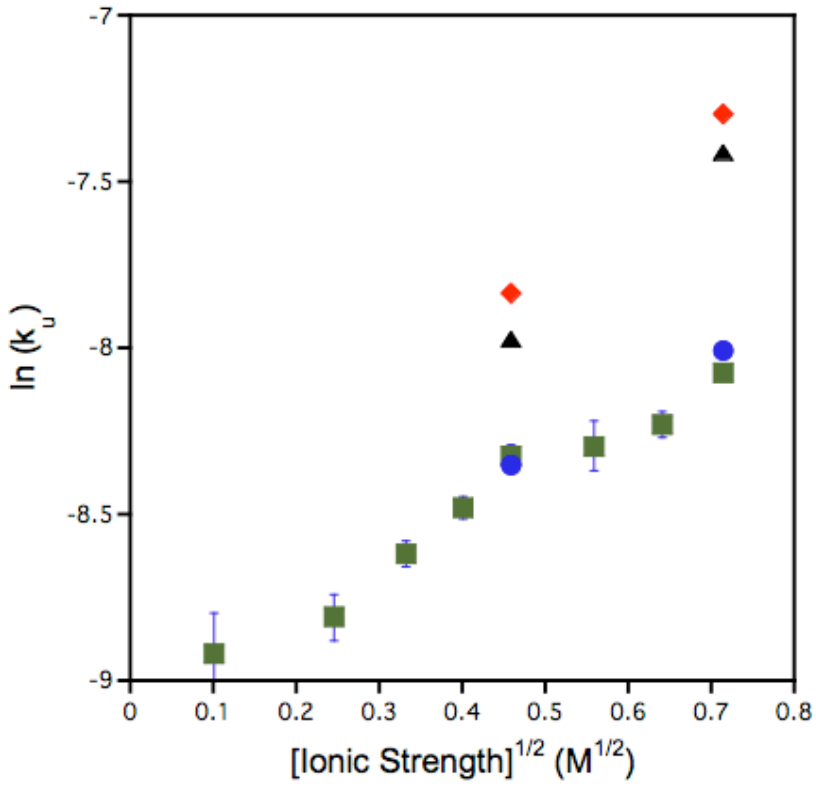


Figure 5.5: The effect of different salts on αLP unfolding at pH 5.
NaCl is shown in green squares, KCl is blue circles, NaBr is red diamonds, and KBr is black triangles. The increased sensitivity to salts with bromide anion suggests specific binding.

To further examine the specific binding of ions by the αLP TS, we measured the unfolding rate dependence of Relocation1 mutant. Specific ion binding is most likely occurring at charged residues, so if the Relocation mutant ameliorates the ion binding effect, then this gives localized information about the TS. However, the Relocation1 mutant exhibits no significant change in the salt dependence (Figure 5.6), indicating that the residues of the relocated ion pair (Asn15B➔Asp , Glu32➔Ala32, Arg141➔Ser141) are most likely not involved in the binding of ions in TS.

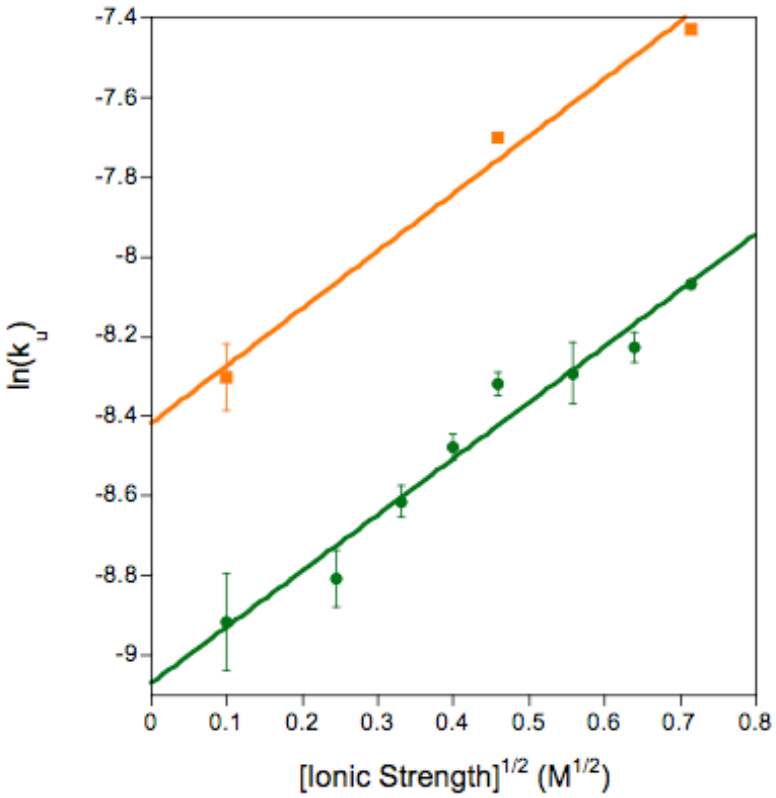


Figure 5.6: The dependence of ionic strength on unfolding of Relocation1 at pH 5.0 is similar to that of αLP.

Although the studies described above have not yet yielded a complete understanding of the electrostatic contribution to kinetic stability, they are quite promising. Further mutagenesis involving relocation and disruption of salt bridges will give highly localized information about the role of electrostatics in attaining kinetic stability. The investigation of specific ion binding can yield useful structural information into the elusive TS. In fact a combination of mutagenesis with pH and ionic strength dependencies will certainly yield the most insight into the electrostatic components that dictate the energetic landscapes for these proteases. Fortunately, Pinar Erciyas will be continuing these promising studies.

Materials and Methods

NAPase was prepared as described(Mitsuiki, Sakai et al. 2002) and lyophilized. Lyophilized protein was reconstituted in water and buffer exchanged into 2mM sodium acetate, pH5. αLP Relocation mutants were created using the Quik Change Multi-Site Mutagenesis. αLP was prepared as described(Mace and Agard 1995). All chemicals were purchased from Sigma-Aldrich.

Unfolding (60 ºC for αLP and the Relocation mutants and 70 ºC for NAPase) was performed using intrinsic tryptophan fluorescence (excitation 283nm, emission 330nm for NAPase, 322nm for αLP and Relocation mutants) using a Fluoromax-3 fluorometer (JY Horiba). The buffers (all at 10mM) utilized were: glycine (pH 2.0, 2.5 and 3.0), potassium acetate (pH 5.0). Unfolding rate constants were obtained by fitting data to a single exponential equation or with a single exponential plus a linear term.

Chapter 6: Mesophile vs. Thermophile: Insights Into the Structural Mechanisms of Kinetic Stability

Preface

To gain insight into the structural basis for kinetic stability, we determined the structure and unfolding behcvior of TFPA, a thermophilic homolog of αLP. From this work we identified a structural element that contributes significantly to the unfolding free energy barrier. These results not only corroborate our earlier model for the unfolding Transition State, but describe a set of detailed interactions that are acting to modulate kinetic stability.

I am responsible for all data and analysis herein. This work has been submitted as an article to the Journal of Molecular Biology.

Abstract

Obtaining detailed knowledge of folding intermediate and transition state (TS) structures is critical for understanding protein folding mechanisms. Comparisons between proteins adapted to survive extreme temperatures with their mesophilic homologs are likely to provide valuable information on the interactions relevant to the unfolding transition. For kinetically-stable proteins such as α-lytic protease (αLP) and its family members, their large free energy barrier to unfolding is central to their biological function. To gain new insights into the mechanisms that underlie kinetic stability, we have determined the structure and high temperature unfolding kinetics of a thermophilic homolog, *Thermobifida fusca* Protease A (TFPA). These studies led to the

identification a specific structural element bridging the N- and C-terminal domains of the protease (the 'domain bridge') that is associated with the enhanced high temperature kinetic stability in TFPA. Mutagenesis experiments exchanging the TFPA domain bridge into αLP validate this hypothesis and illustrate key structural details that contribute to TFPA's increased kinetic thermostability. These results lead to an updated model for the unfolding transition state structure for this important class of proteases in which domain bridge undocking and unfolding occurs at or before the TS. The domain bridge appears to be a structural element that can modulate the degree of kinetic stability of the different members of this class of proteases.

Introduction

Kinetically stable proteins retain their native structure through exceedingly slow unfolding kinetics rather than through a thermodynamic equilibrium favoring the native state (N) over the unfolded state (U).(Baker and Agard 1994) Kinetic stability is increasingly viewed as an important aspect of protein structure and folding, because many proteins (such as capsid protein SHP(Forrer, Chang et al. 2004), lipase(Rodriguez-Larrea, Minning et al. 2006), pyrrolidone carboxyl peptidase,(Kaushik, Ogasahara et al. 2002) subtilisin BPN',(Eder, Rheinnecker et al. 1993) α-lytic protease (αLP),(Baker, Sohl et al. 1992; Sohl, Jaswal et al. 1998) and its homologs *Streptomyces griseus* Protease B (SGPB)(Truhlar, Cunningham et al. 2004) and *Nocardiopsis alba* Protease (NAPase; manuscript submitted)) exhibit the exceedingly slow unfolding kinetics that are characteristic of kinetic stability. Moreover, the concepts that underlie kinetic stability

have important functional implications for a broad array of proteins, regardless of their actual mechanism of native state stabilization. For example, the rate of unfolding has been found to be a determinant of protein evolution, even in proteins whose stability is dictated by thermodynamics, such as thioredoxin.(Godoy-Ruiz, Ariza et al. 2006) In addition, the rate of unfolding is an important factor in the function of some key cellular proteins such as p53,(Butler and Loh 2005; Butler and Loh 2006) and the formation of amyloid fibers,(Canet, Sunde et al. 1999; Thirumalai, Klimov et al. 2003; Ohnishi and Takano 2004; Johnson, Wiseman et al. 2005) ordered protein aggregates associated with various diseases.

The best-studied kinetically-stable proteins are the members of the αLP subfamily of serine proteases. These proteases have extremely large free energy barriers to both folding ($t_{1/2}$ ~ 1800 years for αLP(Sohl, Jaswal et al. 1998), and ~3 days for SGPB(Truhlar, Cunningham et al. 2004)) and unfolding ($t_{1/2}$ ~ 1.2 years for αLP(Sohl, Jaswal et al. 1998) and 11 days for SGPB(Truhlar, Cunningham et al. 2004)). In the case of αLP, the folding landscape is such that the native state is actually less stable than the unfolded state by 4 kcal/mol.(Sohl, Jaswal et al. 1998) To facilitate folding on a biologically relevant time scale, these proteins are synthesized with a covalently-attached pro region, which catalyzes folding of the protease domain by accelerating the folding reaction by factors in excess of 10^9.(Peters, Shiau et al. 1998; Sohl, Jaswal et al. 1998) Once the protease domain has folded, the pro region is proteolytically removed to yield the free, intact protease, effectively decoupling the folding and unfolding landscapes.(Cunningham and Agard 2004) This decoupling has been shown to be crucial for the development of several key functional properties of the αLP native state(Jaswal,

Truhlar et al. 2005) such as its remarkable structural rigidity and resistance to proteolysis.(Jaswal, Sohl et al. 2002)

Studies of αLP and its homologs have provided a broad range of insights into the advantages, limitations, and mechanisms of kinetic stability. Comparison of the αLP and SGPB folding landscapes illustrated that the benefits of increasing the kinetic stability of a protease comes at the cost of requiring a larger Pro region.(Truhlar 2004) In more recent work, a combination of kinetic, structural, and mutational studies of αLP and the acid-resistant homolog NAPase demonstrated the utility of kinetic stability in the evolution of proteins that can survive in a broad range of harsh environments.(manuscript submitted) The αLP-NAPase comparison also led to a structural model of the unfolding transition state (TS) for this class of proteases in which the N- and C-terminal domains separate from each other while remaining relatively intact.

While these comparative studies yielded general insight into the nature of the TS for these proteases, a detailed understanding of how kinetic stability can be modulated and what structural features are responsible for this control remains unknown. Therefore, we sought to gain insight into the structural basis for kinetic stability by comparing the mesophilic αLP to a thermophilic homolog. We identified *Thermobifida fusca* Protease A (TFPA) as a thermophilic αLP homolog from a BLAST search(Altschul, Gish et al. 1990) of the NCBI sequence database (http://www.ncbi.nlm.nih.gov/BLAST/). TFPA is a secreted protease from the organism *T. fusca*(Gusek and Kinsella 1987; Kristjansson and Kinsella 1990) (formerly *Thermomonospora fusca*) that normally grows in decaying plant matter at ~55 ºC. TFPA has high sequence homology to both αLP and NAPase (48.7% and 58.4% identity to each, respectively; Figure 6.1) and is synthesized with a

large (152 residue) and homologous pro region, strongly suggesting that TFPA is kinetically stable as well.

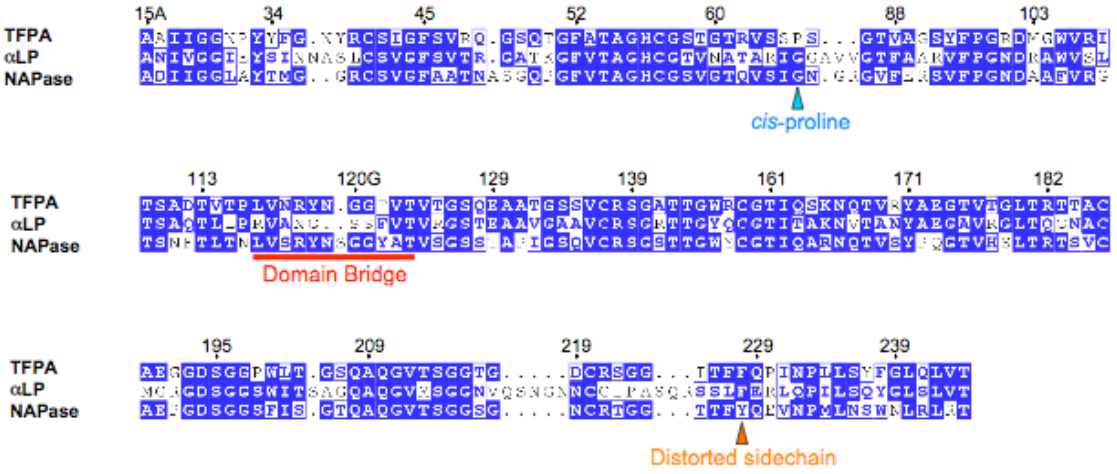


Figure 6.1: Aligned Sequences of TFPA, αLP and NAPase.
Identities are shown in filled blue boxes with white lettering, while similarities are in hollow blue boxes with blue lettering. Sequence alignments(Gouet, Courcelle et al. 1999) are based on structural superposition.(Shindyalov and Bourne 1998) TFPA residues are numbered based on homology to chymotrypsin according to the scheme of Fujinaga *et al.*(Fujinaga, Delbaere et al. 1985) Residues underlined in red comprise the domain bridge, while positions marked with cyan and orange triangles denote the *cis*-proline and distorted phenylalanine residues, respectively.

Previous work indicated that TFPA is quite resistant to both heat and strong denaturants(Kristjansson and Kinsella 1990). Here we show that TFPA is kinetically thermostable; i.e. it attains its remarkable thermostability through extremely slow unfolding kinetics. Our structural analysis helped us identify a specific region of the protein, a β-hairpin connecting the two domains of the protease, which we hypothesize contributes significantly to TFPA's increased kinetic thermostability. Mutational studies and comparisons of homologs confirm this hypothesis and illustrate that this structural element contributes to kinetic stability throughout the entire class of kinetically stable

proteases, thus providing more insight into the TS structure for these proteins. Finally, other possible sources of increased kinetic thermostability are identified and their implications are discussed.

Results

TFPA Unfolds Slowly at High Temperature

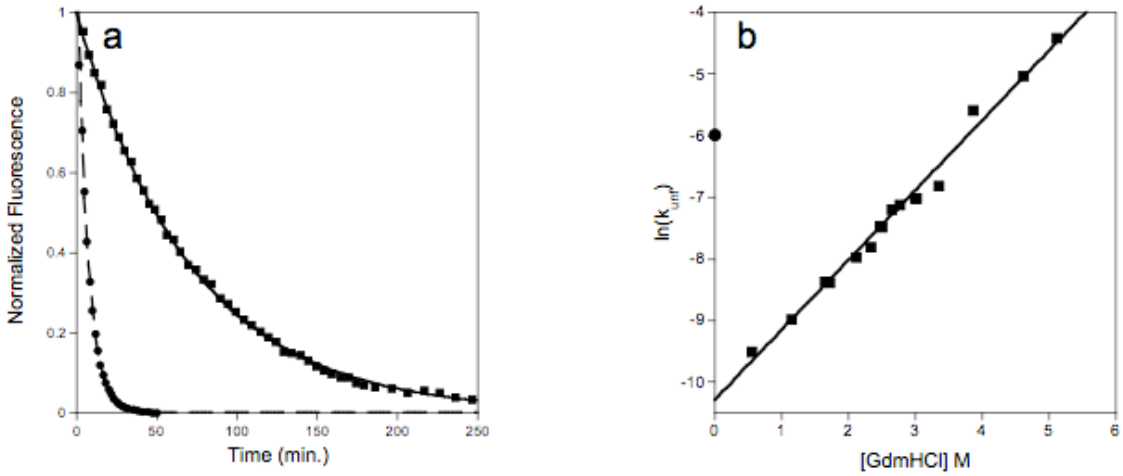


Figure 6.2: Unfolding of TFPA at 70 ºC.
(a) Direct comparison of unfolding of αLP without denaturant (circles, dashed line) and TFPA in 1.72M GdmHCl (squares, solid line) followed by intrinsic tryptophan fluorescence. (b) TFPA unfolding rate determined by linear extrapolation. The unfolding of TFPA (squares; $t_{1/2}$ = 5.5h) is ~70-fold slower than that of αLP (circle; $t_{1/2}$ = 4.5 min) at the same temperature.

Since TFPA is from a thermophilic bacterium and must be adapted to retain longevity under these destabilizing conditions, we expected it to display slow unfolding kinetics, especially at high temperature. TFPA unfolding is readily monitored by measuring intrinsic tryptophan fluorescence (excitation 283nm, emission 330nm) (Figure 2a). Unlike αLP, accurate determination of the TFPA unfolding rate constant at 70 ºC

required denaturant to accelerate unfolding. The $t_{1/2}$ for αLP unfolding at pH 5.0 and 70 ºC is 4.5 minutes (Figure 2), while the unfolding rate for TFPA under identical conditions is ~70-fold slower, as determined by linear extrapolation ($t_{1/2}$ ~ 5.5 hours; Figure 2b). Therefore, TFPA attains its thermal stability through extraordinarily slow unfolding kinetics; it is kinetically thermostable.

Structure of TFPA

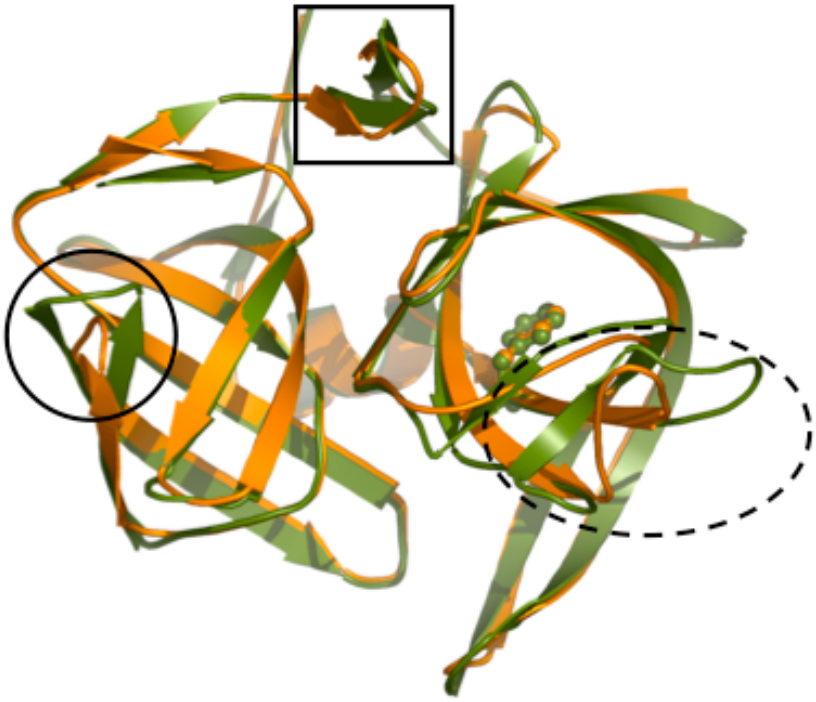


Figure 6.3: Structural superposition of TFPA (orange) and αLP (green) structures. Although the two proteins display remarkable structural similarity (C_α RMSD = 1.1 Å), there are three main regions that show significant differences: loops in the N-terminal domain (circle) and C-terminal domain (oval), and the domain bridge (boxed). Phe228 is shown as ball-and-stick.

To understand the structural basis for TFPA's greater kinetic thermostability, we determined the structure of TFPA by X-ray crystallography to a resolution of 1.44Å (Table 6.1). There were two essentially identical (C_α RMSD = 0.2Å) copies of TFPA per

asymmetric unit in the crystal and, with the exception of the sidechains of the surface residues Gln49 and Arg62 (residues are numbered based on homology to chymotrypsin according to the scheme of Fujinaga *et al*.(Fujinaga, Delbaere et al. 1985)), all atoms could be easily modeled into the electron density. Like all members of the chymotrypsin fold, TFPA has a canonical double β-barrel structure (Figure 6.3), with the active site comprising residues from both the N- and C-terminal domains. As has been observed in other Pro-dependent protease structures(James, Sielecki et al. 1980; Nienaber, Breddam et al. 1993; Kitadokoro, Tsuzuki et al. 1994; Huang, Lu et al. 1995; Fuhrmann, Kelch et al. 2004), TFPA differs from the classical chymotrypsin fold within two structural elements: (1) an 11-residue β-hairpin connecting the N- and C-terminal domains dubbed the 'domain bridge' and (2) a longer, 13-residue β-hairpin within the C-terminal domain which is important for folding catalysis.(Peters, Shiau et al. 1998; Sauter, Mau et al. 1998; Truhlar and Agard 2005)

Table 6.1: TFPA Structure Statistics

Data Statistics	
Space group	C2
Unit cell dimensions (Å)	a=131.96, b=68.55, c=45.36, α=γ=90, β=101.98°
Molecules per asu	2
Limiting resolution (Å)	1.44
I/σ(I)	15.5 (2.6)[a]
Completeness (%)	96.9 (87.8)[a]
R_{merge}[b](%)	0.072 (0.352)[a]
Structure Refinement	
Resolution range (Å)	50 – 1.44
R (%)	20.2 (28.4)
R_{free} (%)	22.9 (33.1)
Protein residues[c]	372
Waters[c]	325
AEBSF ligands[c]	2
Sulfate ions[c]	2
Glycerol[c]	2
Average isotropic B-factors	
Protein atoms (Å^2)[c]	6.5 (±3.1)
Sidechain atoms (Å^2)[c]	7.4 (±3.9)
Solvent atoms (Å^2)[c]	21.6 (±13.7)

[a]Shown in parentheses for the highest resolution bin, 1.47 to 1.44Å
[b]R_{merge} as calculated by Scalepack(Otwinowski 1997)
[c] Values given for the asymmetric unit, which contains two TFPA molecules

Despite excellent overall stereochemistry (Table 6.1), the sidechain of Phe228 is significantly distorted from planarity by 5.5° in molecule A (Figure 6.4) and 4.6° in molecule B of the asymmetric unit. In addition, there was significant distortion of the C_α-C_β-C_γ bond angle in both of these residues (C_α-C_β-C_γ bond angle = 116.6° and 116.4° in molA and molB, respectively.) The distortion of Phe228 is not an artifact of disorder

in the sidechain as the B-factors for this residue are ~50% lower than average for all sidechains. Phe228 is the only sidechain significantly distorted from ideal geometry (3.4 and 2.8σ from the mean for all Phe and Tyr residues). The sidechain distortion was surprising because energetic restraints were used in the refinement process. When the final cycles of refinement were performed without planarity and C_α-C_β-C_γ bond angle restraints, the observed distortion increased dramatically (9.1° and 7.2° planarity distortions, and 119.3 and 119.4° C_α-C_β-C_γ bond angles for mol A and mol B, respectively).

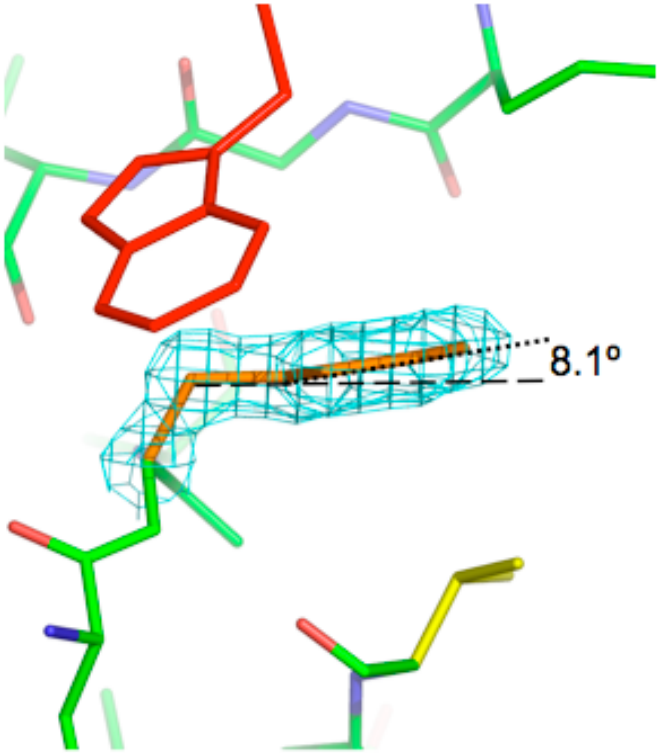


Figure 6.4: Distortion of a buried phenylalanine sidechain.
The phenyl ring of Phe228 in TFPA is deformed from the normal planar conformation by 8.1° due a close contact with Thr181. This distortion was observed in other homologs of TFPA as well, although the distortion angle is greater in TFPA. The model shown is the average coordinates of superposed Phe228 from molecule A and B.

Distortion of this sidechain has been observed previously in the structures of two homologous proteins: (1) 5.8º for Phe228 of αLP at 0.83Å (determined without restraints)(Fuhrmann, Kelch et al. 2004) and (2) 4.4º for Tyr228 of NAPase at 1.85Å (determined with restraints; manuscript submitted).

TFPA exhibits substantial structural homology to αLP (C_α RMSD = 1.1Å; Figure 6.3) and NAPase (C_α RMSD = 0.9Å; Supplemental Figure 6.1), as expected from the extensive sequence homology. Despite this global similarity, three regions of TFPA significantly deviate from those in αLP (Figure 6.3). (1) The largest deviation is in a loop in the C-terminal domain (residues 216 to 226) which is known to help define the protease substrate specificity in αLP.(Mace, Wilk et al. 1995) This loop is much larger in αLP (19 residues), while in all other Pro-dependent proteases, including TFPA, this loop is much shorter (11-residues) and more conserved. Because the longer loop is unique to αLP, it is highly unlikely that these differences are key to the kinetic thermostability of TFPA. (2) Another region that diverges from αLP is a 3-residue loop in the N-terminal domain (residues 66 to 84). In αLP and most other homologs, this loop is much larger and extends outward from the protein core (Figure 6.5b). However, in TFPA, this same loop is mediated by a *cis* proline (Pro 67), which allows for a much tighter turn, thereby shortening the loop by three residues (Figure 6.5a). Additionally, this turn in TFPA is positioned with the proline sidechain pointed into the N-terminal domain hydrophobic core, whereas the turn in αLP is oriented away from the protein interior.

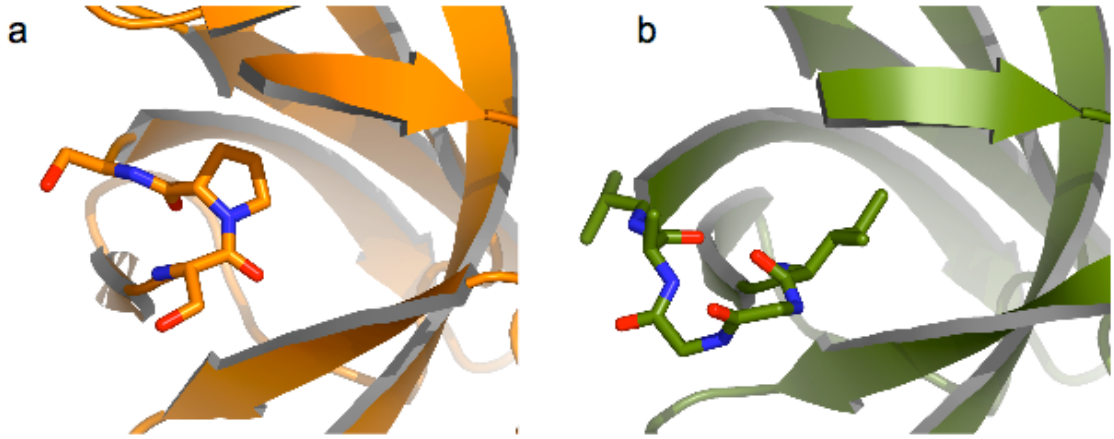


Figure 6.5: TFPA turn mediated by a cis-proline.
(a) The loop in TFPA consist of only three residues, with the turn mediated by a cis-proline. (b) The same loop in αLP is longer, with several residues pointed toward solvent. Two of these residues have the highest B-factors in the protein (Gly 67 and 81).

(3) In TFPA, the domain bridge, one of the structural motifs that are unique to the Pro-dependent proteases (see above), is significantly different in sequence (Figure 6.1) and structure (Figure 6.6a) from that of αLP (Figure 6.6b). In particular, TFPA's domain bridge contains a striking cation-pi (Arg120D-Trp157) interaction that is absent in αLP (Asn120D-Tyr157). The TFPA Domain Bridge most closely resembles that of NAPase, a homologous protease that was also found to be more kinetically thermostable than αLP (Supplemental Figure 6.2; manuscript submitted).

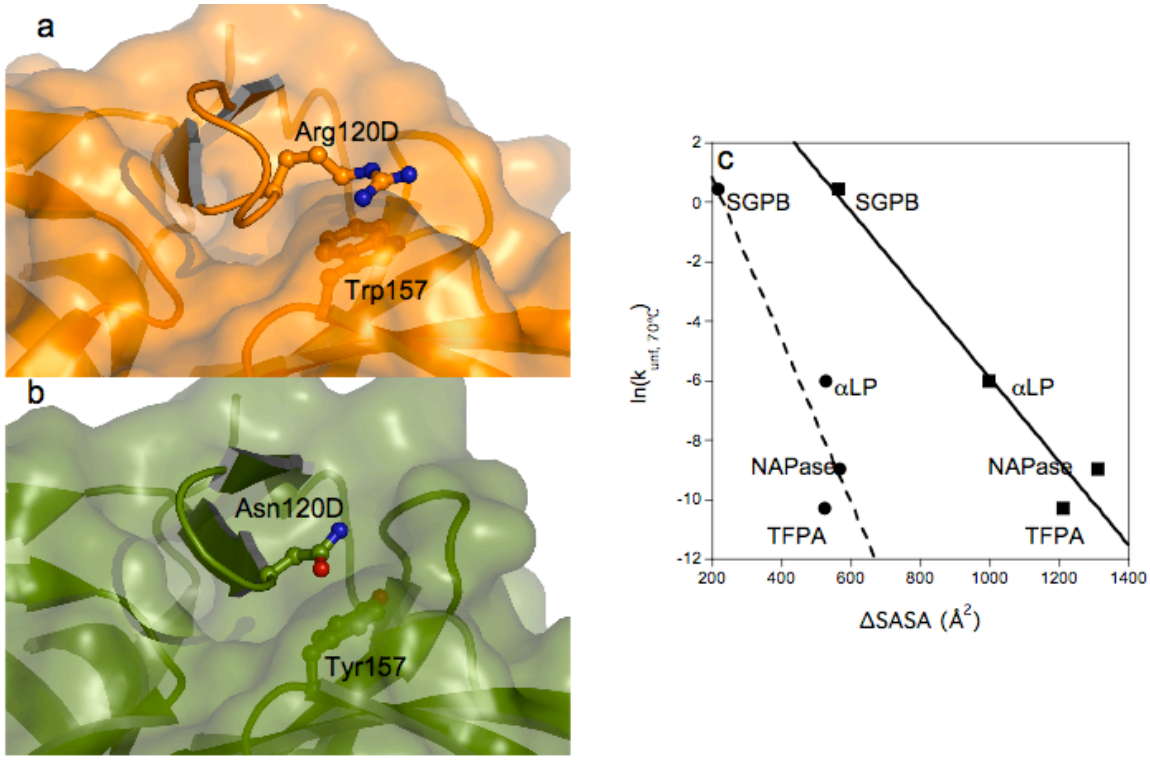


Figure 6.6: Domain Bridge structural attributes are associated with kinetic thermostability.
(a) The domain bridge of TFPA (orange) adopts a different conformation than that of αLP (green). In addition, TFPA contains a cation-pi interaction from Arg120D of the domain bridge to Trp157 of the C-terminal domain which is absent in αLP. (b) The total surface area buried by folding the domain bridge (within the hairpin as well as with the N- and C-terminal domains) strongly correlates strongly ($R^2 = 0.94$; m = 9.5 ± 1.7 cal/mol•K) the height of the unfolding barrier at 70 ºC for the Pro-dependent proteases (squares, solid line). If the calculation is performed with just the hydrophobic surface (circles, dashed line), the slope of this correlation (19 ± 5 cal/mol•K; $R^2 = 0.86$) is similar to that of a Van der Waals interaction (22 cal/mol•K).

Investigating the Role of the Domain Bridge in Kinetic Stability

The provocative differences in domain bridge architecture prompted us to further examine the role of the domain bridge in the unfolding for this class of proteases. Currently, there are four proteases in this sub-family for which both unfolding and structural data are available: (1) SGPB(Huang, Lu et al. 1995; Truhlar 2004), (2) αLP,(Fuhrmann, Kelch et al. 2004; Jaswal, Truhlar et al. 2005) (3) NAPase(manuscript

submitted), and (4) TFPA. Strikingly, there is a remarkable correlation between the unfolding free energy barrier observed at high temperature and the amount of total surface area buried by the folding of the domain bridge and its docking onto the N- and C-terminal domains (Figure 6.6c). This correlation is quite robust ($R^2 = 0.95$) and extends over almost five orders of magnitude in unfolding rate. When just the hydrophobic surface area change is considered, the correlation remains strong ($R^2 = 0.86$) and the slope is 19 (±5) cal/mol•Å^2, suggesting that domain bridge contributions to the unfolding energetics are dominated by hydrophobic interactions (~20 cal/mol•Å^2).(Eriksson, Baase et al. 1992)

Because structural attributes of the domain bridge correlate with kinetic thermostability, we wanted to experimentally delineate the roles of the domain bridge and the cation-pi interaction in kinetic thermostability. Although the domain bridges of αLP and TFPA differ in structure, they dock into a structurally well-conserved pocket formed at the domain interface (Figure 6.6A & B), suggesting that the TFPA domain bridge could be grafted onto αLP without significant steric clashes. Likewise, it would appear that αLP Tyr157 could be changed to Trp to recapitulate the Arg120D interaction in TFPA without generating steric clashes with other structural elements. Therefore, we created a chimeric protease by replacing αLP's domain bridge with that of TFPA (residues 120A thru 120K; Figure 6.1) and mutated Tyr157 to Trp to emulate the cation-pi interaction. Hereafter, we call this mutant BridgeSwap/Y157W. To individually test the contributions of the domain bridge and the cation-pi interaction, we also made separate variants that contained either the BridgeSwap or the Y157W mutations. For each mutant, we again measured the unfolding rate at high temperature (70 ºC).

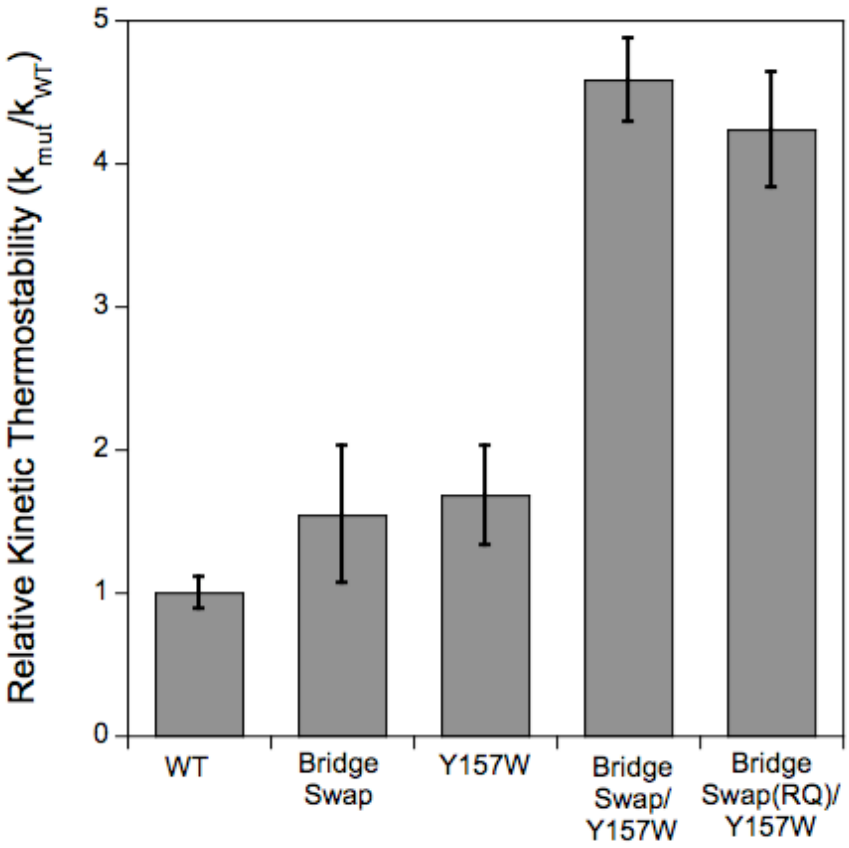


Figure 6.7: Grafting the TFPA domain bridge onto αLP increases its kinetic thermostability.
Shown is the relative effect on the unfolding rate at 70 ºC of different mutants. Only the combination of the domain bridge swap and the Y157W mutant create significantly increased thermostability.

Both the BridgeSwap and Y157W mutants exhibited roughly 1.5 times slower unfolding than wild-type αLP (Figure 6.7). However, when both of these mutations were combined in the BridgeSwap/Y157W variant, there was a >4.5-fold reduction of the unfolding rate. This large, synergistic effect upon creation of the cation-pi interaction suggested that this interaction might play a key role in generating kinetic thermostability. To test this hypothesis, the cation-pi interaction was removed from BridgeSwap/Y157W

by substituting Arg120D with Gln. This variant (hereafter, BridgeSwap(RQ)/Y157W) maintains the packing interactions in the context of the swapped domain bridge, but lacks the cation-pi interaction. Surprisingly, BridgeSwap(RQ)/Y157W, like BridgeSwap/Y157W, unfolds ~4.5-fold slower than WT, illustrating that the cation-pi interaction is, in fact, not important for TFPA's kinetic thermostability. Instead, it appears that the packing interactions between the domain bridge and the N- and C-terminal domains determine the magnitude of the free energy unfolding barrier.

Discussion

The goal of this work is to gain insight into the structural basis for the large unfolding free energy barrier in the α-lytic protease class of Pro-dependent proteases. We expected that the thermophilic homolog TFPA would have a larger unfolding free energy barrier than αLP in order to survive in its extremely harsh environment, and that we could use a structural comparison to pinpoint structural features responsible for its enhanced thermostability. Indeed, TFPA clearly has greater kinetic thermostability than αLP, as exhibited by its slower unfolding rate at high temperature (Figure 6.2). We therefore solved the structure of TFPA to high resolution. From this structure we identified three potential structural determinants for kinetic thermostability (Figure 6.3).

Stabilization Mediated by a *cis*-Proline

The first candidate region is a three residue loop in the TFPA N-terminal domain (residues 66 to 84) whose turn is mediated by a *cis*-proline, allowing the loop to be especially short (Figure 6.5a) and limiting the conformational flexibility of the turn

residues. In contrast, the corresponding loop in αLP is longer by three residues, with a Gly-Gly motif (Figure 6.5B) that is considerably more dynamic in the native state than the rest of the protein, as measured by crystallographic B-factors.(Fuhrmann, Kelch et al. 2004) By mediating a tight turn and minimizing conformational flexibility, the *cis*-proline motif of TFPA could increase the kinetic thermostability. Indeed, proline residues have been shown to increase thermostability in other systems,(Matthews, Nicholson et al. 1987; Nakamura, Tanaka et al. 1997; Muslin, Clark et al. 2002) especially when found in turn positions.(Watanabe, Chishiro et al. 1991; Hardy, Vriend et al. 1993; Watanabe, Masuda et al. 1994; Zhu, Xu et al. 1999) Intriguingly, the only related protease to show significant sequence homology to TFPA in this loop (including a proline at residue 67) is also from an extreme thermophile, *Pyroccocus furiosus* (data not shown).

A Conserved Sidechain Distortion

Another potential contributor to the enhanced kinetic thermostability of TFPA is a phenylalanine residue completely buried in the C-terminal domain. We observed that this residue's phenyl ring was distorted from planarity by a remarkable ~5° with planarity restraints and ~8° without restraints. Significant distortion of Phe228 has also been seen in both αLP(Fuhrmann, Kelch et al. 2004) and NAPase (manuscript submitted). The conservation of sidechain distortion in the same residue in multiple proteins strongly suggests that this distortion has some functional relevance, possibly in generating kinetic stability. In support of this hypothesis, the proteases synthesized with a small pro region, which display a lower degree of kinetic stability(Truhlar, Cunningham et al. 2004), have

Phe228 in a different, unstrained rotamer.(Huang, Lu et al. 1995) The distortion of Phe228 in αLP, determined without energetic restraints imposed in crystallographic refinement, was predicted to have an energetic penalty of ~4 kcal/mol.(Fuhrmann, Kelch et al. 2004) Therefore, it was hypothesized that the distortion could be contributing to αLP's metastability and/or high folding/unfolding free energy barriers.(Fuhrmann, Kelch et al. 2004) Interestingly, the distortion in TFPA is much greater than that found in αLP when the planarity restraints are removed. This suggests that there may be a quantitative correlation between amount of strain at Phe228 and the degree of kinetic stability.

The Domain Bridge Plays a Key Role in Kinetic Thermostability

The domain bridge of αLP is a prime candidate for modulating kinetic stability for three reasons. First, the entire structural element is unique to this class of Pro-dependent proteases and significantly differs between αLP and TFPA, both in terms of sequence (Figure 6.1) and its detailed interactions (Figure 6.6A & B). Secondly, the TFPA domain bridge most closely resembles that of NAPase (Supplemental Figure 6.2), a homologous protease that also exhibits greater kinetic thermostability than αLP (manuscript submitted). Finally, a remarkable correlation between domain bridge interaction surface area and the height of the unfolding free energy barrier (Figure 6.6c) indicates that the domain bridge may be playing a role in kinetic stability across the entire subfamily of pro-dependent proteases. Could the domain bridge and its interactions account for at least some of TFPA's increased thermostability?

To test the role of the TFPA domain bridge in generating kinetic thermostability, we made chimeric mutants that recapitulate the interactions of the TFPA domain bridge

within αLP. By increasing the domain bridge interaction surface area as found in TFPA (the BridgeSwap/Y157W variant), unfolding can be slowed by >4.5-fold (Figure 6.7), indicating that these interactions play a key role in determining the unfolding barrier height. This is the first time we have been able to increase the unfolding free energy barrier in αLP, indicating that we are beginning to understand the structural mechanisms underlying kinetic stability. In fact, when the unfolding barriers of SGPB, αLP, NAPase, and the BridgeSwap/Y157W mutant are compared as a function of the domain bridge length (a surrogate for the domain bridge interaction area since there is no structure of BridgeSwap/Y157W) a virtually perfect correlation is obtained (Supplemental Figure 6.3; $R^2 = 0.99$). The slope indicates that the rate of unfolding slows ~5-fold for every residue added into the domain bridge. This suggests that the BridgeSwap/Y157W mutant exhibits the full kinetic stabilization expected from the TFPA domain bridge, indicating that there must be other regions in TFPA that contribute to its larger unfolding barrier (see above.) These results illustrate the utility of thermophilic and mesophilic comparisons and thermophilic/mesophilic chimeras to gain insight into important issues of protein stability and folding.

In agreement with the results of the TFPA/αLP chimeras, there is other evidence that the domain bridge is a key component of the unfolding transition. First, our lab observed ~4-fold faster unfolding at high temperature (E. Cunningham and DAA, unpublished data) with an αLP mutant in which the domain bridge has been extended by 8 residues by inserting a TEV protease recognition site(Cunningham and Agard 2003). This extension would be expected to decrease domain bridge stability due to increased loop entropy because the extended residues should be unstructured. The mutation lowers

the unfolding barrier by 0.9 (±0.2) kcal/mol, in reasonable agreement with loop entropy estimates ($\Delta\Delta G = -1.5RT\ln(L/L_o) = 0.6$ kcal/mol, where L_o and L are the original and the new loop lengths, respectively; Ref. Grantcharova *etal*(Grantcharova, Riddle et al. 2000)) Second, high temperature (500K) Molecular Dynamics simulations of wt-αLP unfolding indicate that interactions along the domain bridge are preferentially broken in the unfolding transition state (N. Salimi and DAA, unpublished data). Finally, a genetic screen designed to isolate αLP mutants with increased resistance to high temperature identified multiple, distinct mutations in the domain bridge (BAK & DAA, unpublished data).

The results presented here also give insight into the nature of the interactions that the domain bridge utilizes to enhance kinetic stability. The observed correlations between the unfolding free energy barrier and either hydrophobic surface area (Figure 6.6c) or number of contact residues (Supplemental Figure 6.3) and the lack of a correlation with the cation-pi interaction (also supported by the NAPase structure in which the cation-pi interaction could occur but doesn't, Supplemental Figure 6.2) collectively indicates that the domain bridge modulates kinetic stability exclusively via its packing interactions with the rest of the protease.

Not only does the domain bridge clearly play an important role in dictating the unfolding free energy barrier for TFPA and αLP, but there is also evidence that the domain bridge plays a role in kinetic stability across all proteases of this class (Figure 6.6c). This insight provides further detail into a developing structural model of the unfolding TS for these proteins. In a recent study (manuscript submitted), we showed that the interactions along the domain interface are broken in the TS for these proteases,

which would necessitate the involvement of the domain bridge in the unfolding process. The results described here not only support this conclusion, but also indicate a more crucial role for the domain bridge than had been previously appreciated. Although other models for the unfolding TS are possible, we favor one in which the domain bridge undocks from the N- and C-terminal domains while also losing some of its structure (Figure 6.8). Since the unfolding free energy barrier depends linearly on the surface area buried by the domain bridge (Figure 6.6c), these interactions must be broken as well during the unfolding TS (Figure 6.8).

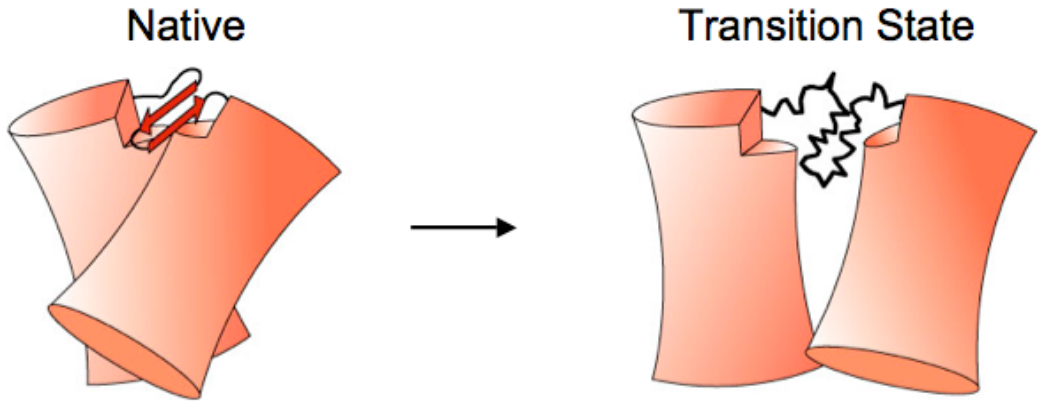


Figure 6.8: Model for the unfolding transition state for the Pro-mediated proteases. In the TS, the N- and C-terminal domains separate, yet remain intact themselves [Kelch etal, submitted], while the domain bridge breaks many key contacts holding these two domains together.

In addition, the domain bridge β-hairpin must lose some structure in the TS as well, as loop entropy of the domain bridge effects the total unfolding rate. Therefore, the domain bridge could be thought of as a linchpin holding the two domains together, thereby slowing unfolding and extending the protease lifetime(Jaswal, Sohl et al. 2002). Additionally, the domain bridge is a structural feature whose interaction surface has been expanded and optimized throughout evolution to increase the kinetic stability of these

proteases. This would suggest that the interface of the N- and C-terminal domains is already highly optimized across this protease family and that the domain bridge acts as an adjustable element that can modulate the degree of kinetic stability of each family member.

Materials And Methods

Protein Expression and Purification

TFPA was heterologously produced in the S164 strain of *Streptomyces lividans* which contains the pGL-1 plasmid for constitutive expression of TFPA.(Lao and Wilson 1996) A starter culture of 100mL Tryptic Soy Broth with 50μg/L of the antibiotic thiostrepton was inoculated with 0.2mL of an S164 glycerol stock and grown overnight at 30 ºC. 20mL of this starter culture was added per 1L YZZ media (16g NZ amine, 5g yeast extract, 2.5g NaCl per L) supplemented with 5mM HEPES, pH 7.3, 1g/L agar, 0.05g/L EDTA, 0.2g/L $MgSO_4$, 8μg/L $ZnSO_4$, 20μg/L $FeSO_4 \bullet 7H_2O$, 15μg/L $MnSO_4 \bullet 7H_2O$, 26μg/L $CaCl_2 \bullet 2H_2O$ and 50μg/L thiostrepton. Steel springs were placed around the interior of the 2L flask to break up the conglomerates of bacteria for maximal expression yields. This culture was shaken at 150 rpm and 33 ºC for 5 days.

Cultures were centrifiuged at 3000 x g for 30 minutes and the cell pellet discarded. The culture supernatant was heated to 70 ºC for 4 hours, which removed a majority of contaminating proteins. The culture supernatant was diluted 1:10 with cold 10mM sodium acetate, pH 5.0 and the final pH was adjusted to ~4.7 with 1N acetic acid. 5mL of SP-sepharose per mg of TFPA was added and stirred overnight at 4 ºC to bind. The beads were allowed to settle and the supernatant discarded. The beads were added to

a column (Biorad, Hercules CA) and washed with 5 column volumes of 10mM sodium acetate, pH 5.0, followed by a wash of 5 column volumes of 10mM glycine, pH 8.4. Pure TFPA was eluted with 2 column volumes of 300mM NaCl, 10mM glycine, pH 8.4. TFPA was buffer exchanged to remove residual salt and concentrated in a Biomax 5kDa MWCO filter (Millipore) that has been pretreated with poly-lysine to limit TFPA binding to the membrane.

The BridgeSwap variant of αLP was created using recursive PCR(Prodromou and Pearl 1992) with primers 5'-AACCGCTACAACGGCGGAACGGTCACCGTGCGCGGC-3' and 5'-CGTTCCGCCGTTGTAGCGGTTCACGCGCGGCAGCAG-3' using the pALP12 vector as the template.(Mace, Wilk et al. 1995) The BridgeSwap/Y159W and Y159W variants were made with QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) using primers 5'-GATGGTGCCGCACTGCCAACCGGTGGTGCGGCC-3' and 5'-GGCCGCACCACCGGTTGGCAGTGCGGCACCATC-3' and BridgeSwap and the pALP12 plasmid(Mace, Wilk et al. 1995) as templates, respectively. The BridgeSwap(RQ)/Y159W variant was also created using QuikChange with primers 5'-CTGCCGCGCGTGAACCAGTACAACGGC-3' and 5'-GCCGTTGTACTGGTTCACGCGCGGCAG-3' and the BridgeSwap/Y159W variant as template. Mutant proteins were expressed and purified as previously described.(Mace, Wilk et al. 1995) All chemicals were purchased from Sigma Aldrich, except GdmHCl (ICN Biochemicals).

Protein Unfolding

TFPA unfolding at 70 ºC was measured by fluorescence (excitation =283nm, emission = 330nm) using a Fluoromax-3 fluorometer (J. Y. Horiba) connected to an external water bath. Unfolding was initiated by manual mixing of concentrated TFPA into preheated 10mM potassium acetate (pH 5.0) to a final [TFPA] = 0.1-1 μM. The unfolding rate constant was obtained by fitting the data to either a single exponential equation or a single exponential with a linear term. Denaturant concentration was measured directly using the refractive index of the solution.(Pace 1986) Unfolding of αLP variants were similarly measured except the fluorescence emission wavelength was 322nm. The SGPB unfolding rate at 70 ºC was extrapolated from Jaswal *et al.*(Jaswal, Truhlar et al. 2005) $\Delta G^{\ddagger}_{N\text{-}TS}$ values derived from Figure 6.6c and supplemental Figure 6.3 were calculated using standard Transition State theory with a "pre-exponential factor" of k_BT/h. Although the appropriate pre-exponential factor for an unfolding or folding reaction is controversial,(Dimitriadis, Drysdale et al. 2004; Kubelka, Hofrichter et al. 2004) the correlation and slope are unaffected by the choice of a pre-exponential term.

TFPA Crystallizaiton and Structure Determination

Crystallization was accomplished using the hanging-drop, vapor diffusion method in 24-well trays (Nextal Bitechnologies, Montreal, Canada). Crystals of apo-TFPA were grown with 1:1 drops of 10mg/ml protein and 8% (w/v) PEG8000, 20mM $NiCl_2$, 0.1M Tris, pH 8.5. Subsequent crystallizations were set up using TFPA inhibited with the serine protease inhibitor aminoethylbenzylsulfonyl fluoride (AEBSF). To generate inhibited protein, 100 molar excess AEBSF in 100mM HEPES, pH 7.0 was added to concentrated TFPA and this solution was incubated at room temperature overnight. Near

100% inhibition was obtained. TFPA-AEBSF was then concentrated in a BioMax 5 kDa molecular weight cutoff filter and buffer exchanged with 2mM sodium acetate, pH 5.0 in the concentrator. Large, diffraction-quality crystals of TFPA-AEBSF (4mg/ml) were obtained in 8% (v/v) glycerol, 10mM $MnCl_2$, 0.15M $AmSO_4$, 0.1M Tris pH 8.5.

Diffraction data were collected at the Advanced Light Source (Beamline 8.2.1 for apo crystals and 8.2.2 for inhibited crystals) and processed using HKL2000.(Otwinowski 1997) Apo crystals (space group $P2_1$) diffracted to 2.1Å with three TFPA molecules per asymmetric unit. Initial phases were obtained using molecular replacement in CNS(Brunger, Adams et al. 1998) using a poly-Ala model of αLP, with non-homologous loops removed to minimize model bias. An apo model was manually built in O(Jones 1990) and refined in CNS,(Brunger, Adams et al. 1998) with all backbone atoms and most sidechains modeled.

TFPA-AEBSF crystals (space group C2, with 2 TFPA molecules per asymmetric unit) diffracted to much higher resolution (1.44Å), although these crystals suffered from very high and anisotropic mosaicity (0.6-1.9). Phases were obtained using molecular replacement in CNS(Brunger, Adams et al. 1998) with an apo-TFPA search model. Initial rounds of refinement were performed in CNS(Brunger, Adams et al. 1998) using NCS averaging, and subsequently performed in Refmac(Murshudov, Vagin et al. 1997), using TLS refinement(Winn, Isupov et al. 2001) with each TFPA monomer as an independent unit (and no NCS averaging). Electron density maps were high quality and nearly all sidechains could be unambiguously modeled except for Gln49 and Arg62 (in both TFPA molecules in the asymmetric unit.). According to the structure validation program PROCHECK(Laskowski, Macarthur et al. 1993), the resulting TFPA-AEBSF

model had acceptable or better geometry with 90.2% of residues are in the most favourable conformation, while the remaining 9.8% are in additionally allowed regions of Ramachandran space.

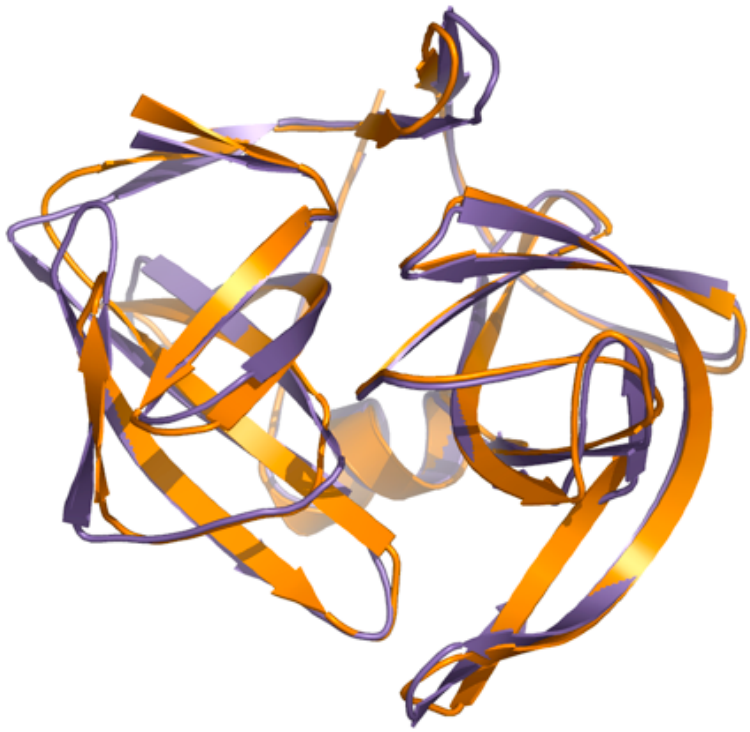
Structural Analysis

Total surface areas were calculated in CNS(Brunger, Adams et al. 1998) using a probe volume of 1.6Å and hydrophobic surface area calculated using the Conditional Hydrophobic Accessible Surface Area algorithm(Fleming, Fitzkee et al. 2005)(http://roselab.jhu.edu/chasa/runchasa.html). To calculate the surface area of the N- and C-terminal domains buried by the domain bridge, the domain bridge (residues 120A through 120K) was treated as a separate protein and the surface area at the interface was measured. The area buried upon folding of the domain bridge β-hairpin was measured by treating each strand as separate polypeptides (in isolation from the rest of the protease). To correct for the artificial surface area component arising from the 'cleaved' peptide bond, the surface area along each 'cleavage site' was determined separately and subtracted from the values given above.

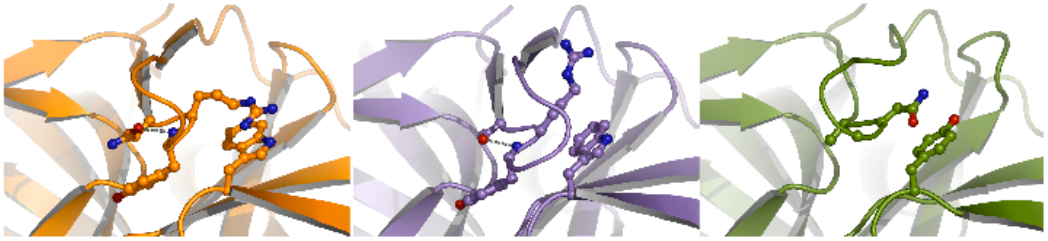
Structural alignments were performed using the combinatorial extension method.(Shindyalov and Bourne 1998) These structure-based alignments were also used to create the sequence alignments of Figure 6.1.(Gouet, Courcelle et al. 1999) Figures 6.3, 6.4, 6.5, and 6.6, and Supplemental Figures 6.1 and 6.2 were made in Pymol.(DeLano 2002)

Acknowledgements

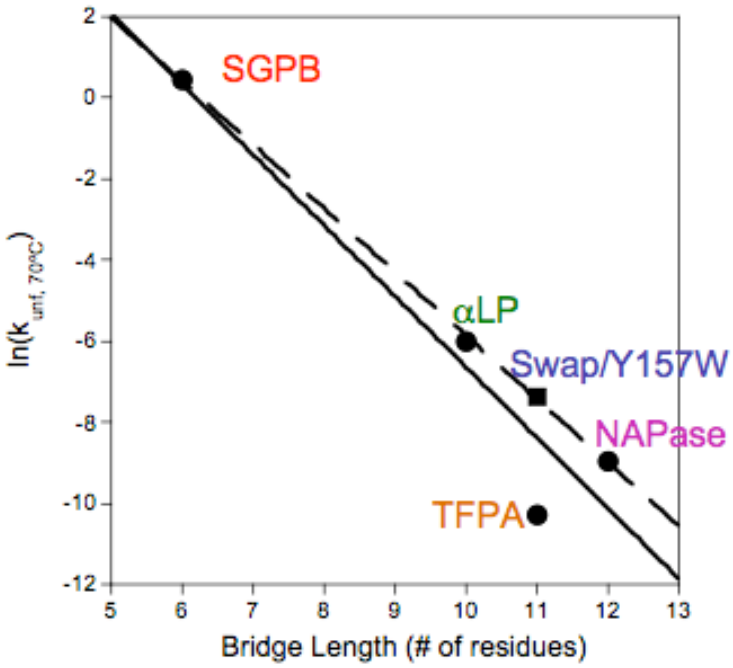
We would like to thank Diane Irwin and Dr. David Wilson at Cornell University for the generous gift of the S164 strain of *S. lividans* and for helpful advice on expression and purification. We greatly appreciate Drs. C. Fuhrmann and L. Rice for assistance with diffraction data collection, processing and refinement. We are grateful to A. Baucom for assistance with Figure 8 and Dr. G. Murshudov for assistance with Refmac. We'd also like to thank Q. Justman, Drs C. Fuhrmann, S. Truhlar and L. Rice for critical reading of the manuscript and members of the Agard lab for helpful suggestions. BAK was supported by an HHMI Predoctoral Fellowship.



Supplemental Figure 6.1: Structural homology between TFPA and NAPase.
The overall structure of TFPA (orange) is extremely similar to that of NAPase (violet) (C_α RMSD = 0.9 Å).



Supplemental Figure 6.2: Structural Similarity of the TFPA and NAPase domain bridges. The TFPA (orange) and NAPase (violet) domain bridges are similar in the backbone conformation as well as the sidechain interactions. Conversely, the αLP domain bridge (green) adopts a different architecture and has very different interactions with the N- and C-terminal domains.



Supplemental Figure 6.3: Comparison of Domain Bridge length and unfolding free energy of activation.
The solid line is a linear fit to the four Pro-dependent proteases SGPB, αLP, NAPase and TFPA ($R^2 = 0.92$). A linear fit to the same data, but replacing TFPA with the αLP mutant Swap/Y157W, yields a much better correlation (dashed line; $R^2 = 0.99$)

Chapter 7: Detailed Characterization of TFPA Unfolding

Preface

TFPA is from a thermophilic organism and, therefore, unfolds much slower than αLP at high temperature. To understand the structural basis for TFPA's enlarged unfolding barrier, I attempted to perform Eyring Analysis to determine the quasi-thermodynamic parameters that govern the TFPA unfolding free energy barrier in a manner analogous to that performed for αLP and SGPB. However, the extremely slow unfolding kinetics prevented accurate determination of unfolding rates at low enough temperature to use the Eyring Equation. In lieu of this, I analyzed the TFPA unfolding data using the Arrhenius equation, which provides some insights into the structural basis for the increased unfolding barrier and agrees well with the structural model proposed in the previous Chapter. In addition, I wanted to examine the role of cooperativity in the unfolding free energy barrier of TFPA. Gladiator experiments using TFPA and other proteases indicate that TFPA may not be as ultra-cooperative as other homologs such as αLP and SGPB.

Sheila Jaswal performed all the unfolding of αLP that I used to compare with TFPA. I am responsible for all TFPA data and analysis.

Introduction

The unfolding barrier is key to the longevity of kinetically stable proteins, because the large unfolding barrier serves to maintain the protein in the native state (Baker, Sohl et al. 1992; Baker, Shiau et al. 1993; Sohl, Jaswal et al. 1998). However,

the height of the unfolding barrier is not the only aspect of energy landscape that is key to protein longevity. The unfolding cooperativity (i.e. the ruggedness of the unfolding free energy barrier) also can be crucial for protein longevity, especially in a highly proteolytic environment(Park and Marqusee 2004; Park and Marqusee 2005) and this aspect of the landscape has been optimized as well in kinetically stable proteins(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). This is not necessarily true of thermodynamically stable proteins, such as trypsin, which has an unfolding barrier height comparable to that of αLP but is degraded ~20-fold faster(Truhlar, Cunningham et al. 2004). Therefore, the energetics of the unfolding barrier, both in terms of its height and its cooperativity, are crucial for establishing protein longevity. In αLP and SGPB, decoupling of the folding reaction from the unfolding reaction appears to be key for the development of these energetic characteristics(Jaswal, Truhlar et al. 2005). Although the unfolding barrier is clearly a crucial aspect of the function of αLP and its homologs, the structural basis for its characteristics remain largely unknown.

By correlating differences in unfolding properties with structural attributes between homologs, the mechanistic underpinnings of kinetic stability could become clear. Therefore, we studied the detailed characteristics of the unfolding free energy barrier of a thermophilic αLP homolog, *Thermobifida fusca* Protease A or TFPA(Lao and Wilson 1996). As expected for a thermophilic protein, TFPA unfolds significantly slower than the mesophilic αLP, especially at elevated temperature Chapter 6. To identify the detailed energetics that underlie this perturbation in the TFPA unfolding barrier, we attempted to measure the thermodynamics parameters that constitute the

unfolding free energy barrier in TFPA using Eyring Analysis. In addition, we used proteolysis to probe the cooperativity of the TFPA unfolding barrier.

Results and Discussion

Eyring Analysis of TFPA

To perform Eyring Analysis, we measured the TFPA unfolding rate over a large range in temperature. From this data, the thermodynamic parameters of the unfolding free energy barrier can be determined using the Eyring equation:

$$\ln(k_U/T) = (\Delta S^{\ddagger}_{T0,U} - \Delta C_p{}^{\ddagger}_{T0,U})/R + \ln(h/k_B) + (\Delta C_p{}^{\ddagger}_{T0,U} - (\Delta H^{\ddagger}_{T0,U}/T_0)/R) * (T_0/T) - (\Delta C_p{}^{\ddagger}_{T0,U}/R)*\ln(T_0/T)$$

where R is the gas constant, k_B is Boltzmann's constant, h is Planck's constant, $\Delta H^{\ddagger}_{unf}$ is the enthalpy of activation, $\Delta S^{\ddagger}_{unf}$ is the entropy of activation, and $\Delta C_p{}^{\ddagger}_{unf}$ is the change in heat capacity of the free energy barrier.

A severe difficulty arises directly from the same fact that makes TFPA's unfolding barrier so interesting: folding is so slow that accurate measurement of a rate constant becomes extremely difficult and time-consuming. Direct measurement of the TFPA unfolding rate can only be realistically achieved at extremely high temperature (>70 ºC). Therefore, unfolding rate constants at only three temperatures (75, 80, and 85 ºC) could be measured directly (Figure 7.1). For some of these high temperature experiments, the unfolding rate was variable so multiple unfolding time courses were performed and the results averaged. In addition, each time course was performed with a different TFPA concentration, which allows us to test whether autolysis is affecting the

unfolding. This doesn't seem to be the case because there is no increase in unfolding rate with increasing TFPA concentration.

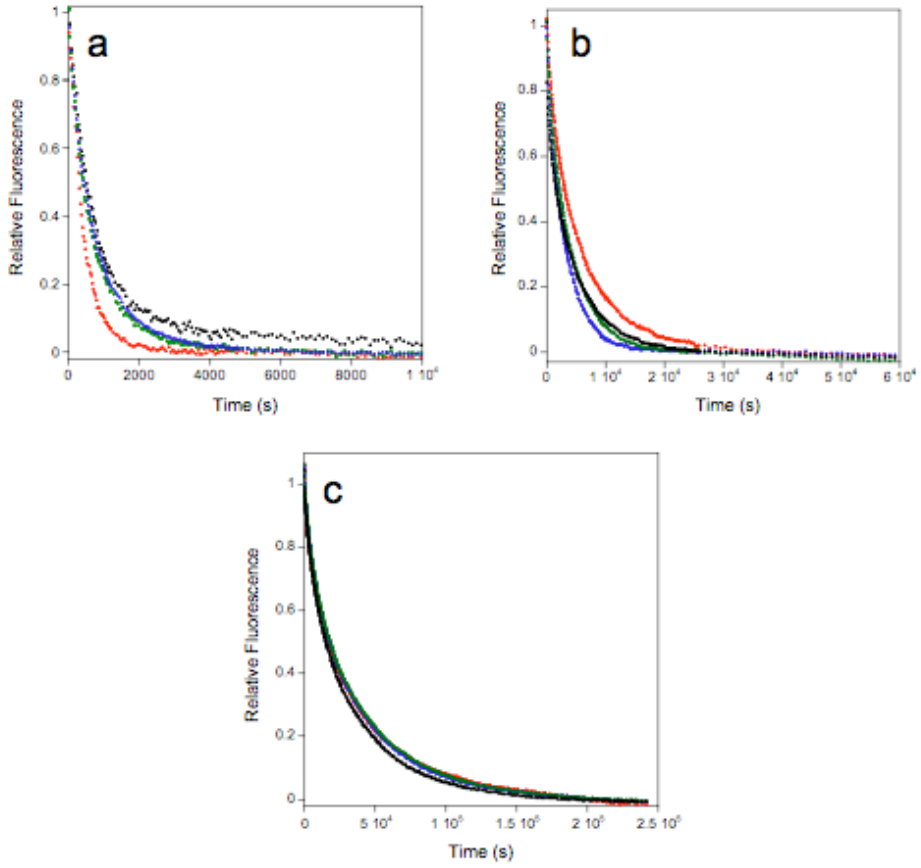


Figure 7.1: Direct unfolding of TFPA
A) Direct unfolding of TFPA at 85 ºC. Unfolding was measured with various TFPA concentrations: 50nM (black), 100nM (red and green), 200nM (blue). The aggregate unfolding rate is 1.8 (±0.5) x 10^{-3} s^{-1}. B) Direct unfolding of TFPA at 80 ºC. Unfolding was measured with various TFPA concentrations: 100nM (red and blue), 200nM (green), 300nM (black). The aggregate unfolding rate is 2.4 (±0.5) x 10^{-4} s^{-1}. C) Direct unfolding of TFPA at 75 ºC. Unfolding was measured with various TFPA concentrations: 50nM (red), 100nM (blue and green), 200nM (black). The aggregate unfolding rate is 4.6 (±0.2) x 10^{-5} s^{-1}.

At 70 ºC and lower, it was necessary to use a strong denaturant, such as guanidinium hydrochloride, to accelerate unfolding. An accurate rate constant could be

determined by plotting the unfolding rate against guanidinium concentration and extrapolating to zero denaturant. The Linear Extrapolation model provided excellent fits to the data at moderately high temperature (45-70 ºC; Figure 7.2). This was a bit surprising because the unfolding rates for both αLP and SGPB showed significant curvature across a large range in temperatures (25-60 ºC)(Jaswal 2000; Jaswal, Truhlar et al. 2005). Although this curvature in the denaturant-dependence of unfolding could be accurately modeled with the Denaturant binding model, the root cause of the curvature was determined to be due a mixture of effects due to interactions of the denaturant with the protein and with solvent, i.e. electrostatic effects(Jaswal 2000). The lack of curvature in TFPA unfolding suggests that the electrostatic contribution to unfolding is not as prevalent in TFPA as in αLP or SGPB. In support of this conclusion, the TFPA domain interface, like that of NAPase, has less ionic character than that of αLP and SGPB, which may make its unfolding less dependent on electrostatic effects.

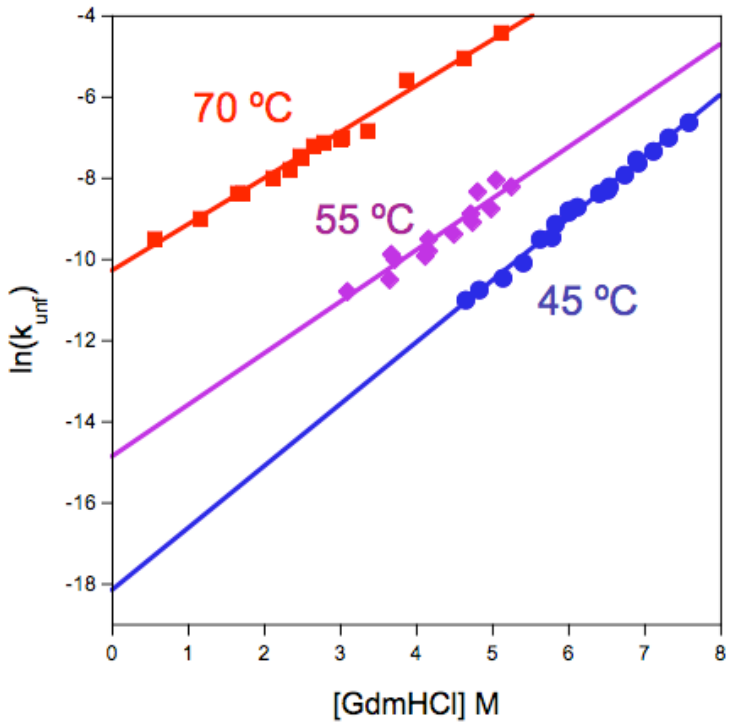


Figure 7.2: High temperature, denaturant titration of TFPA unfolding.
Linear extrapolation of the unfolding barrier to zero denaturant determines the intrinsic unfolding rate constant at 45 ºC (blue; 1.3 (±0.2) x 10^{-8} s^{-1}), 55 ºC (violet; 3.5 (±2.4) x 10^{-7} s^{-1})), and 70 ºC (red; 3.4 (±0.3) x 10^{-5} s^{-1})).

For TFPA, accurate rates ≤37 ºC were extremely difficult to obtain from extrapolation. Because the unfolding rate was so slow, measurement of a large enough span of denaturant for accurate extrapolation was not possible using GdmHCl as denaturant. Therefore, we employed the much stronger denaturant GdmSCN, which contains two highly chaotropic ions: guanidinium cation and thiocyanate anion. Therefore the denaturing activity of GdmSCN is over twice that of GdmHCl.

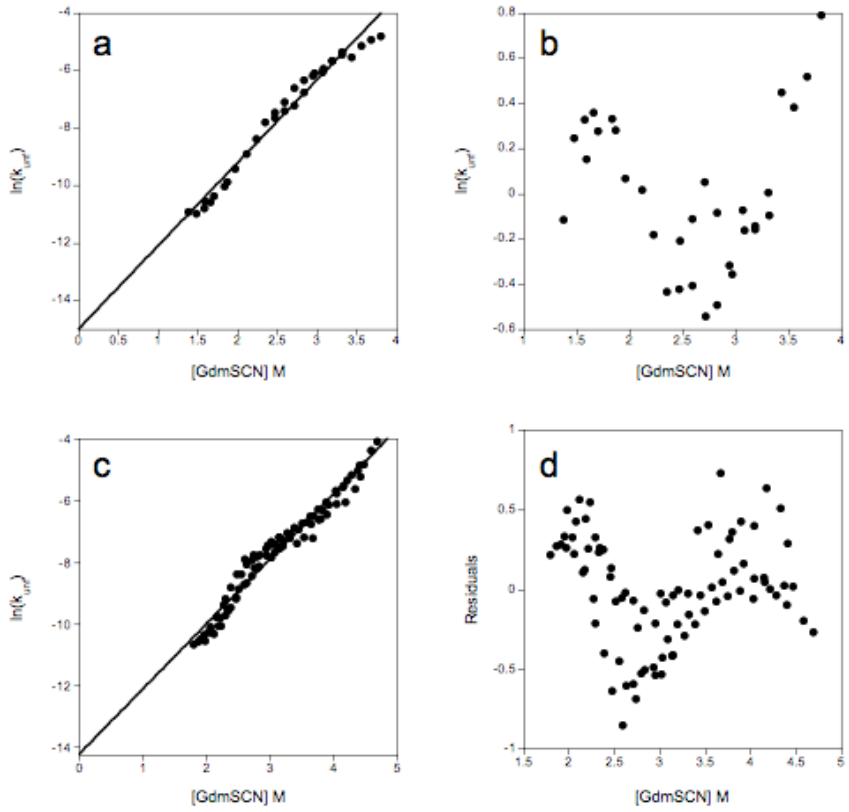


Figure 7.3: Low temperature unfolding of TFPA.
A) TFPA unfolding denaturant titration at 33 ºC. The fit with a linear equation is meant to emphasize the complex curvature in the data. B) Residuals from the linear fit of the 33 ºC data displaying a duck-like shape, further illustrating the complex behavior. C) TFPA unfolding denaturant titration at 20 ºC displaying a nematode-like shape. Once again, the linear fit is meant to emphasize the complex curvature. D) Residuals from the linear fit of the 20 ºC data displaying a peahen-like shape, very unlike the expected random dispersal that would be expected with a fit to the correct model.

While unfolding could be readily measured under a relatively wide concentration of GdmSCN and at various temperatures, the data was highly non-linear (Figure 7.3A and C). In contrast to the monotonic non-linearity observed in αLP and SGPB (Jaswal 2000; Jaswal, Truhlar et al. 2005), the logarithm of TFPA unfolding rate is a complex, antitonic function of GdmSCN concentration. This can be seen even more clearly by observing the residuals of a linear fit which display a complex and non-random behavior (Figure 7.3B and D). We currently do not know what the cause of such behavior is, but it

could be due to specific ion binding to distinct states(Krantz and Sosnick 2001), movement of the Transition State in response to denaturant(Pappenberger, Saudan et al. 2000), light-induced degradation of thiocyanate or, more likely, a mixture of these effects.

To avoid the complications found with the application of GdmSCN, we used GdmHI, which is a more active denaturant than GdmHCl but less so than Gdm SCN. While unfolding could be measured with GdmHI at these temperatures, interpretation of the unfolding kinetics was severely complicated by the oxidation of iodide ion to molecular iodine (I_2) during the time course of the experiment, which could be easily seen by the formation of a brown color, indicative of I_2, over the time frame of minutes to hours. Since the denaturant concentration was changing during the course of the experiment, the unfolding rate constant could not be properly attributed to the correct GdmHI concentration. To combat the rapid oxidation of HI, the reductant Tris[2-carboxyethyl] phosphine (TCEP) was added to the unfolding reaction in relatively high quantities (0.5 M). However, the oxidation of HI to I2 could still be seen (although the oxidation was slowed) and the TCEP could be causing reduction of disulfide bonds, further complicating the interpretation of the unfolding kinetics. For these reasons, this entire approach was abandoned.

Because of the inherent difficulty in attaining unfolding rates at the low temperatures necessary to detect the curvature due to a non-zero $\Delta C_p^{\ddagger}$, the temperature dependence of TFPA unfolding was fit with the linear Arrhenius equation:

$$\ln(k_U/T) = \Delta S^{\ddagger}_{T0,U}/R + \ln(h/k_B) + (\Delta H^{\ddagger}_{T0,U}/RT)$$

, where k_u is the unfolding rate constant, k_B is the Boltzmann constant, h is Planck's constant, R is the gas constant, $\Delta H^{\ddagger}$ is the enthalpy of activation and $\Delta S^{\ddagger}$ is the entropy of activation. Because this analysis assumes a zero $\Delta C_p^{\ddagger}$, we fit the high temperature (70 to 37 ºC) pseudo-linear portion of the αLP data for comparison (Figure 7.4).

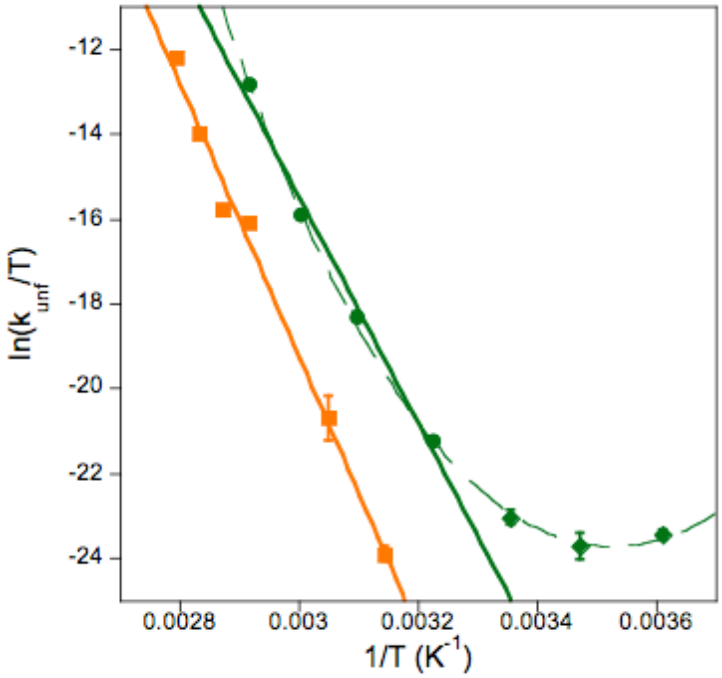


Figure 7.4: Arrhenius analysis of TFPA and αLP unfolding.
TFPA unfolding rates are shown in orange squares. αLP unfolding rates (green) fit with Arrhenius equation are shown as circles, while those left out of the fit are shown as diamonds. The fit of the all aLP unfolding rates to the Eyring equation is displayed as a broken line.

This analysis gives $\Delta H^{\ddagger}_{TFPA}$ = 63.9 (±2.8) kcal/mol, $\Delta S^{\ddagger}_{TFPA}$ = 106 (±8) cal/mol•K, $\Delta H^{\ddagger}_{\alpha LP}$ = 53.0 (±3.9) kcal/mol, and $\Delta S^{\ddagger}_{\alpha LP}$ = 81 (±12) cal/mol•K. This suggests that the TFPA unfolding free energy barrier is more enthalpic in origin at these temperatures than is that of αLP. However, the TFPA unfolding free energy barrier has more stabilizing entropy than that of αLP.

These results coincide nicely with our model for the unfolding model of these proteases (Figure 8, Chapter 6). The increased enthalpic barrier for TFPA unfolding could come from the larger interaction surface of the domain bridge in TFPA than in αLP. Likewise, the slightly larger entropic benefit of TFPA unfolding could come from the larger domain bridge becoming, to some degree, disordered in the TS (Chapter 6; Figure 8).

The Cooperativity of TFPA Unfolding

The cooperativity of unfolding is clearly an important aspect of the energetic landscape and requires optimization for maximal protein longevity(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). The cooperativity of unfolding can be probed using proteolysis because local, as well as global, unfolding can lead to protease accessibility. Indeed, proteolysis has been shown to be a useful tool for examining sub-global fluctuations of proteins in a manner analogous to hydrogen exchange experiments.(Rupley 1967; Park and Marqusee 2004; Park and Marqusee 2005) Proteolytic susceptibility is conveniently determined using a 'survival assay',(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004) in which multiple proteases are combined in equimolar ratios and proteolysis is allowed to proceed. For SGPB and αLP, the rate of degradation matched that of their global unfolding rates, indicating that the unfolding free energy barriers for these proteins are almost perfectly cooperative(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). It is therefore expected that the other homologs would exhibit similar cooperativity. Because αLP, TFPA and trypsin have largely orthogonal substrate specificities (data not shown), one can readily measure the

concentration of active protease remaining over time by measuring the activities of each enzyme using three different chromogenic substrates, each specific for a different protease.

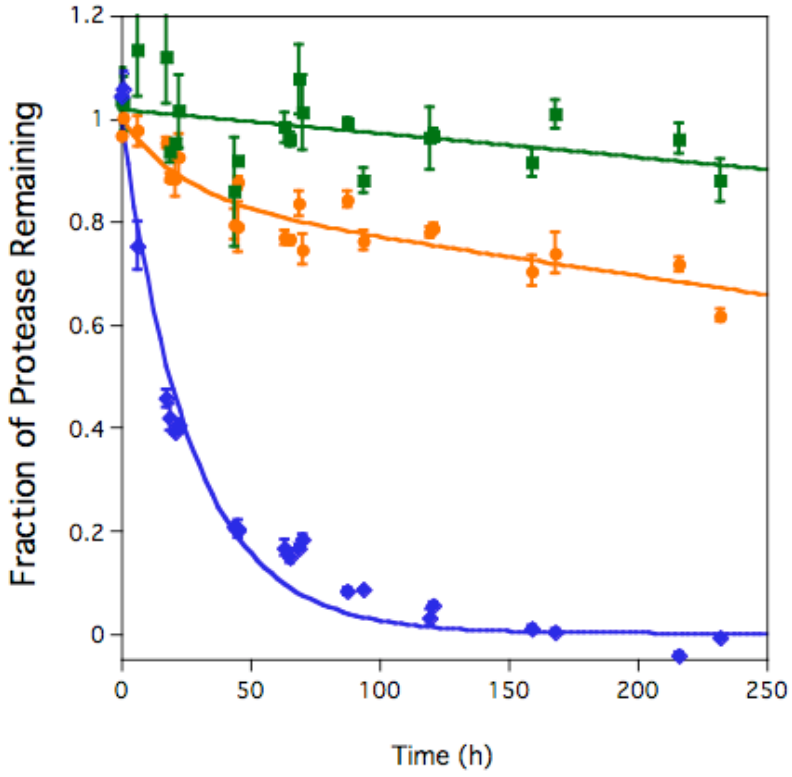


Figure 7.5: Survival Assay of αLP (green), TFPA (orange), and trypsin (blue) at pH 7 and 37 ºC.
αLP data is fit with a linear equation, TFPA with a single exponential plus a linear term and trypsin with a single exponential equation. αLP is degraded ~2-fold slower than TFPA. $k_{deg,\alpha LP}$ = 4.6 (±2.2) x 10^{-4} h^{-1}, $k_{deg1,TFPA}$= 0.040 (±0.0.25) h^{-1}, $k_{deg2,TFPA}$= 7.3 (±2.8) x 10^{-4} h^{-1} $k_{deg,trypsin}$ = 0.037 (±0.002) h^{-1}.

A survival assay was performed using 4μM αLP, TFPA, and trypsin at pH 8.0 and 37 ºC. While trypsin was degraded much faster ($k_{deg,trypsin}$ = 0.037 (±0.002) h^{-1}) than either of the kinetically stable proteases as expected, it was surprising to find that TFPA was actually degraded faster than αLP ($k_{deg,\alpha LP}$ = 4.6 (±2.2) x 10^{-4} h^{-1}; Figure 7.5). Close examination of the TFPA degradation data showed that the loss of activity was biphasic,

with the fast phase ($k_{deg1,TFPA}$= 0.040 (±0.025) h^{-1}) accounting for ~15% total activity loss while the slow phase ($k_{deg2,TFPA}$= 7.3 (±2.8) x 10^{-4} h^{-1}) accounts for the remaining 85%. Although the TFPA unfolding rate has not been measured at 37 ºC (see above), both degradation phases are much faster (~10^7-fold faster than the slow degradation phase) than would be expected based on the linearly extrapolated unfolding rate. This result suggests that TFPA is actually degraded faster than its global unfolding rate. In support of this hypothesis, the fast phase of TFPA degradation corresponds to the portion of the time course when trypsin was still present. If significant sub-global fluctuations occur in TFPA, then degradation would be expected be faster the TFPA in the presence of trypsin.

Another survival assay was performed at 55 ºC (the optimal growth temperature of *T. fusca*), this time using only TFPA and αLP to avoid the complication of trypsin's presence. In this case, TFPA easily outlasted αLP ($k_{deg,\alpha LP}$ = 2.97 (±0.02) x 10^{-5} s^{-1}), but biphasic TFPA degradation kinetics were again observed (Figure 7.6). The fast and slow phases ($k_{deg1,TFPA}$= 4.4 (±0.4) x 10^{-6} s^{-1}, $k_{deg2,TFPA}$= 7.8 (±1.4) x 10^{-8} s^{-1}) both accounted for ~50% of the total activity loss and the slow phase was only ~2-fold faster than the global unfolding rate at 55 ºC measured separately (Figure 7.2). This suggests that the slow phase represents global unfolding and degradation. Once again, the fast phase roughly corresponds to the portion of the time course when αLP is present, which suggests that this phase corresponds to degradation from αLP due to sub-global fluctuations. It appears that TFPA does not have significant autolysis because the inactivation rate is only accelerated when other proteases are present. This suggests that the loops that become sensitive to proteolysis do not contain aromatic character because TFPA is specific for Tyr, Phe and Trp.

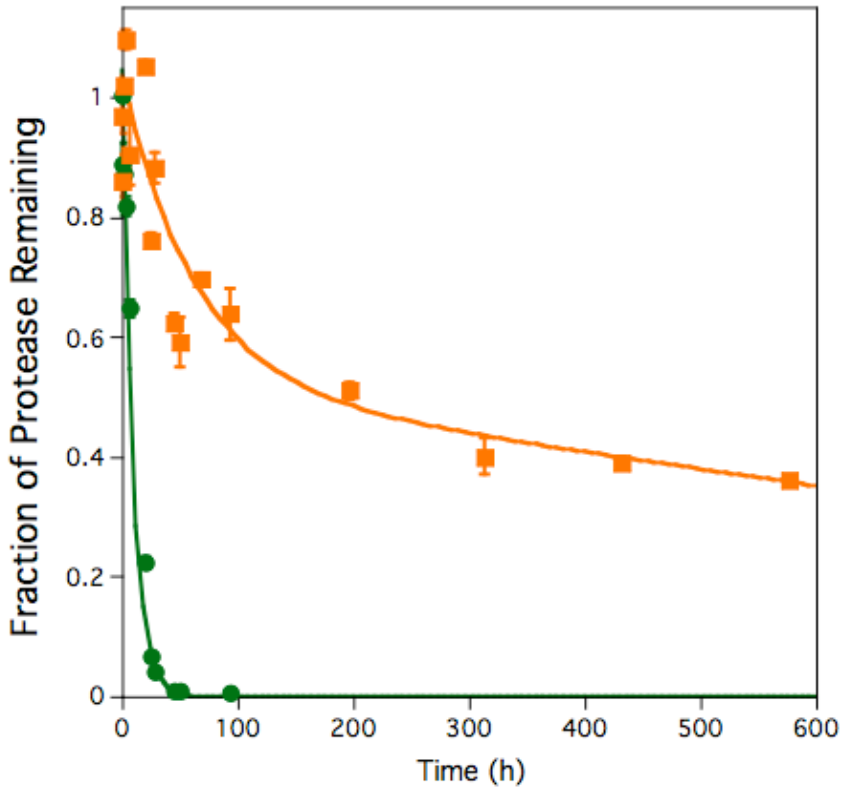


Figure 7.6: Survival Assay of αLP (green) and TFPA (orange) at pH 7 and 55 ºC. αLP data is fit with a single exponential equation whereas TFPA is fit with a single exponential plus a linear term. αLP is degraded much faster than TFPA. $k_{deg,\alpha LP}$ = 2.97 (±0.02) x 10^{-5} s^{-1}, $k_{deg1,TFPA}$= 4.4 (±0.4) x 10^{-6} s^{-1}, $k_{deg2,TFPA}$= 7.8 (±1.4) x 10^{-8} s^{-1}.

This degradation must be due to local unfolding or breathing motions because the global unfolding rate is so much slower than the degradation rates observed in the survival assays, both at high and moderate temperature. What could be the cause of this behavior? This could be due to the sample aging over time. Because TFPA loses a significant amount of activity upon freezing (data not shown), the protein is stored on ice which could allow for oxidation of the protein. This could result in a certain percentage of the protein being oxidatively damaged and, therefore, degraded faster. This putative oxidatively-induced effect would still be due to local unfolding and breathing motions

rather than global unfolding since the global unfolding rate doesn't change as the protein ages (data not shown). Repeating this assay with fresh TFPA could address this issue.

Conversely, the increased sub-global unfolding could be an inherent property of TFPA. If this is true, it raises many questions and opens up many avenues for experimentation. First, this phenomenon could be investigated further by measuring the effect of varying αLP concentration on the amplitude and rate of fast degradation at 55 ºC. One would expect that increasing the starting αLP concentration would increase the rate and amplitude of this phase. This information could be used in a manner analogous to hydrogen exchange to obtain information about the ruggedness of the unfolding free energy barrier. Mass spectrometry could be used in conjunction with the degradation assay to determine the initial cleavage sites to localize the more flexible regions of the protein. Also, one could use the protease substrate specificity to determine what regions in TFPA contain residues consistent with the degradation patterns observed. For example, the higher sensitivity of TFPA to hydrolysis by trypsin than αLP (Figure 7.5) suggests that there is significant cleavage occurring at arginine and/or lysine residues.

Finally, is there a biological reason for the lack of cooperativity in the TFPA unfolding barrier? Is the lack of cooperativity utilized to inactivate the protease by the host bacterium under conditions where active TFPA is not desirable? Conversely, could the lack of cooperativity represent a lack of need for the cooperativity to be maintained throughout evolution. For example, if TFPA does not require resistance to any protease but itself, then may not need evolve a completely cooperative unfolding free energy barrier as found in αLP and SGPB (Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). If TFPA has evolved to be resistant to itself, that may be all that is required for its

survival. For example, *S. griseus* secretes a large number of proteases with varying substrate specificity. Therefore, SGPB requires an ultracooperative unfolding barrier to prevent proteolysis. T. fusca does not seem to secrete any other proteases, which may alleviate the need for extreme cooperativity.

Materials and Methods

TFPA was expressed and purified as described (see Chapter 6). Unfolding was followed by intrinsic tryptophan fluorescence as previously described (Chapter 6). Survival assays were performed as previously described (Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004). Guanidinium hydrochloride was purchased from ICN biochemicals and guanidinium iodide from NIGU Bioselect. All other chemical were purchased from Sigma Aldrich.

Chapter 8: Directed Evolution of αLP to Identify Mutants that Modulate the Folding Landscape

Preface

The structural mechanisms underlying kinetic stability are unknown. Because of the extremely slow kinetics of folding and unfolding, classical folding techniques such as comprehensive Phi-value analysis are not possible to perform on a reasonable timescale. To identify the structural determinants for this behavior, I used directed evolution techniques to identify mutants of αLP that have perturbed energy landscapes. One screen is for mutants that fold faster, while the other is for mutants that unfold slower. These studies, while promising, were abandoned for a more rational approach.

I am responsible for all work described herein.

Introduction

The last several chapters have described approaches that use evolutionary relationships as a means of determining the structural mechanisms underlying kinetic stability. However, these approaches suffer from the fact that natural evolution does not employ one selective pressure to a protein. Therefore, different proteins face many different pressures simultaneously, which makes it very difficult to pinpoint which residues are instrumental in the development of certain behavior.

However, it is possible to determine this information directly in the lab by administering selective pressures on the protein *in vitro* to evolve only the behavior that

is being screened for. The process of directed evolution has been a great asset to understanding protein function in the past (MacBeath, Kast et al. 1998; Pedersen, Otzen et al. 2002; Hocker 2005), and similar techniques have been employed in αLP(Derman and Agard 2000). Random mutants of αLP were screened for their capacity to fold with a Pro Region variant (Pro-3) that is ~1000-times less efficient at catalyzing αLP folding(Peters, Shiau et al. 1998). After screening through over 300,000 colonies, only one mutant of αLP (R138H/G183S) was obtained(Derman and Agard 2000). Remarkalbly, this mutant folds ~300-fold faster than wild-type in both the Pro-catalyzed and uncatalzyed reactions, despite the fact that the rates for these disparate folding reactions differ by a factor of $>10^9$(Derman and Agard 2000). This result suggests that the Transition State structures for the catalyzed and uncatalyzed reactions are extremely similar(Derman and Agard 2000), thus illustrating the usefulness of a random screening approach in the elucidation of key aspects of protein folding.

To address the structural mechanisms by which the folding landscape can be perturbed, this random mutagenesis approach was further extended. Two different screening strategies were employed, one to identify mutants that fold even faster than the R138H/G183S mutant and another to identify mutants that unfold slower than wild-type.

Results/Discussion

A Screen for Faster Folding αLP mutants

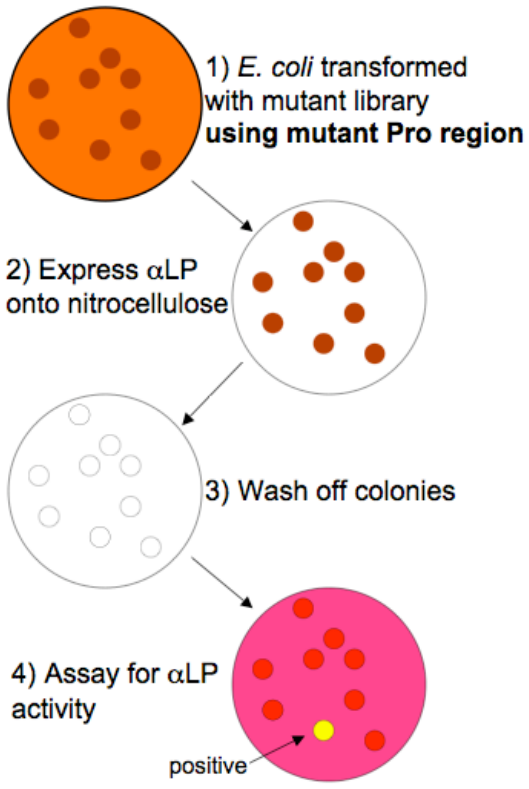


Figure 8.1: Schematic describing the screen for faster folding αLP.

To identify residues that are critical in determining the folding barrier for αLP, I employed a genetic screen for mutants that would lower the folding barrier. The general concept underlying the screen (Figure 8.1) is the same as that of Derman and Agard(Derman and Agard 2000), but the details are different in two key aspects. First, the mutagenesis protocol was greatly improved to expand the breadth of sequence space searched by the screen (see Materials and Methods for details). Second, this study's goal was to generate a series of mutants that constitute a large continuum of kinetic stability to identify interactions that modulate the folding landscape for these proteases. Therefore, this screen, schematized in Figure 8.1, was for mutants that folded faster than the R138H/G183S mutant rather than wild-type. Thus, the gene carrying the R138H/G183S

mutant was subjected to random mutagenesis. In addition, the expression of the mutant library was performed at higher temperature (25 ºC) than in Derman and Agard (room temperature(Mace and Agard 1995)) to decrease the foldase activity of the Pro-3 variant further. At these modestly elevated temperatures, the Pro region is greatly destabilized(Peters, Shiau et al. 1998), thus decreasing the foldase efficiency.

Table 8.1: Mutants obtained from screen for faster folding αLP

Mutant*	# of times obtained	Interesting Interactions
A135V	5	Y26 of the Pro
A202T	2	E30 of the Pro
G18D, G193D	1	Domain Interface
G197D, Q237K	1	Domain Interface

*all in background of R138H/G183S, chymotrypsin-homology numbering scheme

Out of ~10000 colonies screened, nine active clones were obtained. The αLP gene in these clones was sequenced to determine the alterations at the amino acid level (Table 8.1). Two mutations were obtained repeatedly: A135V (5 times) and A202T (twice). These two mutations cluster on the surface of the C-domain along an interface formed between the protease and the Pro Region N-domain(Sauter, Mau et al. 1998) (Figure 8.2A and B). This suggests that these residues play a key role in folding catalysis by the Pro Region. In fact, A135 is in direct contact with Tyr26 of the Pro(Sauter, Mau et al. 1998) (Figure 8.2A), a residue that is known to be crucial for foldase activity(Cunningham, Mau et al. 2002). Mutation of alanine to valine may increase the affinity between αLP and the Pro Region due to increased packing with Tyr26 (Figure 8.2A), which could explain the putative effect on folding of this mutant. In addition, A202 is in the vicinity of Glu30 of the Pro Region(Sauter, Mau et al. 1998) (Figure

8.2B), another residue that is crucial for folding catalysis(Cunningham, Mau et al. 2002). The mutation of Ala202 to threonine could increase the affinity of the Pro Region-protease interaction by creating another hydrogen bond with Glu30 (Figure 8.2B). Despite these plausible explanations for these mutations' effects, none of these mutants has been tested for its role in Pro-mediated folding.

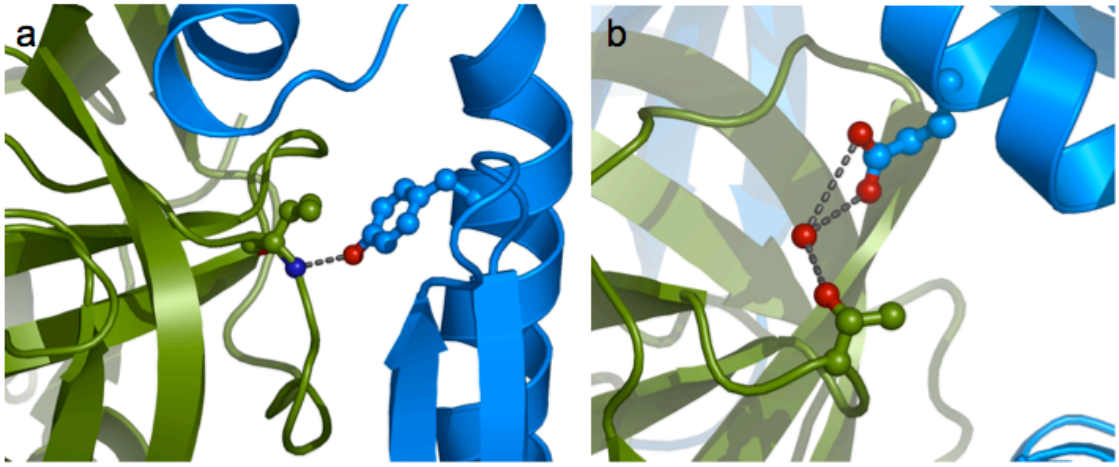


Figure 8.2: Models of the A135V and A202T mutants.
A) Model of the A135V-αLP: Pro Region complex. A135V would allow for additional packing between αLP and Tyr30 of the Pro Region, an important residue for folding catalysis(Cunningham, Mau et al. 2002). B) Model of the A202T-αLP: Pro Region complex. A202T would allow for an additional water-mediated hydrogen bond with Glu26 of the Pro Region, another residue critical for folding catalysis(Cunningham, Mau et al. 2002).

Two other mutants were obtained through this screen: G18D/G193D and G197D/Q237K. One striking observation is the prevalence of glycine to aspartate mutations, which could be due, in part, to the high glycine content of αLP and the ease conversion from a glycine codon to that of aspartate (one G ⇒A transition.) A more subtle observation is that many of the mutations lie along the domain interface (G18D,

G193D, and G197D; Figure 8.3A and B). It has been shown that the domain interface opens during the unfolding TS with the newly created crevice being filled with water molecules (Chapter 4), which leads to potential roles of these mutations in adjusting the folding barrier of αLP. It could be that the hydrophilic aspartate sidechains aid in the solvation of the domain interface, which would stabilize the TS, causing folding to be accelerated. However, neither of these mutants have been shown to have an repeatable effect on the folding barrier nor been experimentally tested *in vitro* for its effect on the folding landscape, nor have their effects been repeatable (unlike A135V and A202T), so these results may be an artifact of the screen.

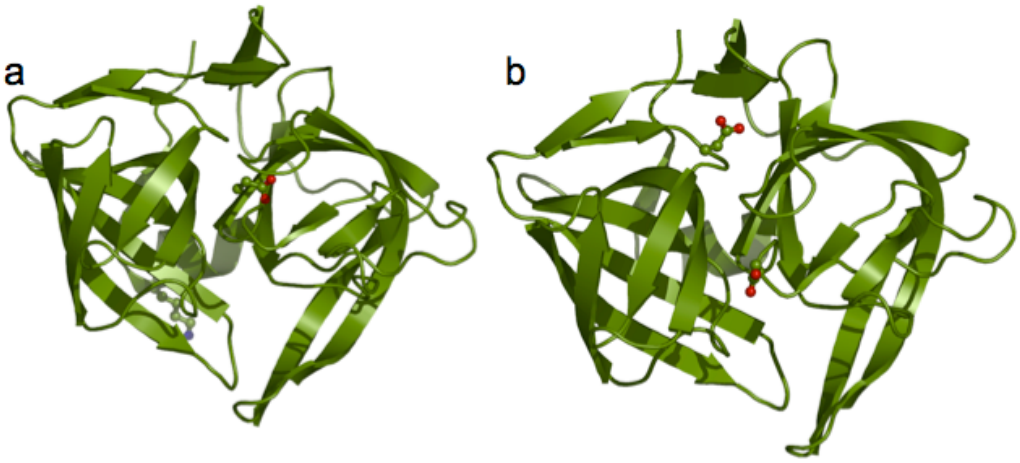


Figure 8.3: Mapping the G145D/Q190K (A) and G18D/G193D (B) mutants onto the structure of αLP.
Mutant residues are shown in ball and stick.

Although the results of this screen illuminate interesting interactions in the Pro-mediated folding pathway of αLP, the major goal was to understand the structural determinants that modulate the uncatalyzed reaction. The organization of the mutants on the surface of αLP and their direct interactions with the Pro region suggest that these

mutations are specifically affecting the catalyzed pathway and not the uncatalyzed. A possible cause for this result is that the selective pressure placed on the protease is not strong enough to identify the interactions that are common to both the catalyzed and uncatalyzed reactions. In addition, destabilization of the Pro may not screen for interactions important for the actual catalysis of folding, but for binding of the intermediate, i.e. mutations that affect K_M rather than k_{cat}.

To address this issue, an alternative selective pressure was employed in the screen. It is known that perturbation of the C-terminal residues of the Pro region decreases foldase activity mainly through lowering the k_{cat} (Peters, Shiau et al. 1998). The Pro-3 variant itself has a ~1000-fold decrease in k_{cat} compared to WT-Pro. However, the Pro-4 variant is so drastically debilitated ($k_{cat, Wt}/ k_{cat, Pro-4} < 10^6$) that it is too radical of a selective pressure to reasonably screen for mutants. Therefore, the final residue in the Pro-3 variant (Leu163) was mutated to alanine to generate a Pro region (Pro-3Ala4) with a folding defect debilitation presumably intermediate between Pro-3 and Pro-4. Indeed, the Pro-3Ala4 variant was incapable of folding the R138H/G183S αLP mutant as measured by plate assays, indicating a folding defect more drastic than Pro-3.

However, no mutants were obtained after screening through ~4000 colonies of a R138H/G183S -αLP mutant library. This failure to identify mutants could be a result of insufficient sampling of the mutant library, as the number of mutant colonies screened is 100-fold less than that of Derman and Agard(Derman and Agard 2000). Alternatively, the failure may be due to the Pro-3Ala4 variant being too debilitated to reasonably obtain αLP mutants that are folding competent. Without an *in vitro* characterization of Pro-3Ala4 catalysis, this remains an unanswered question. Because of these ambiguities, this

approach was dropped in lieu of a more directed and rational mutagenesis strategy (see Chapter 3).

A Screen for Slower Unfolding αLP Mutants

The unfolding barrier is the sole determinant of αLP longevity(Sohl, Jaswal et al. 1998). The entire Pro-folding pathway has been evolved to optimize the unfolding barrier, both in terms of its height but also its cooperativity(Jaswal, Sohl et al. 2002). However, information as to the structural determinants of this barrier remain unknown. Identification of residues that are crucial to the development of this large, cooperative barrier would shed light onto the mechanism by which αLP and its homologs maintain their native states.

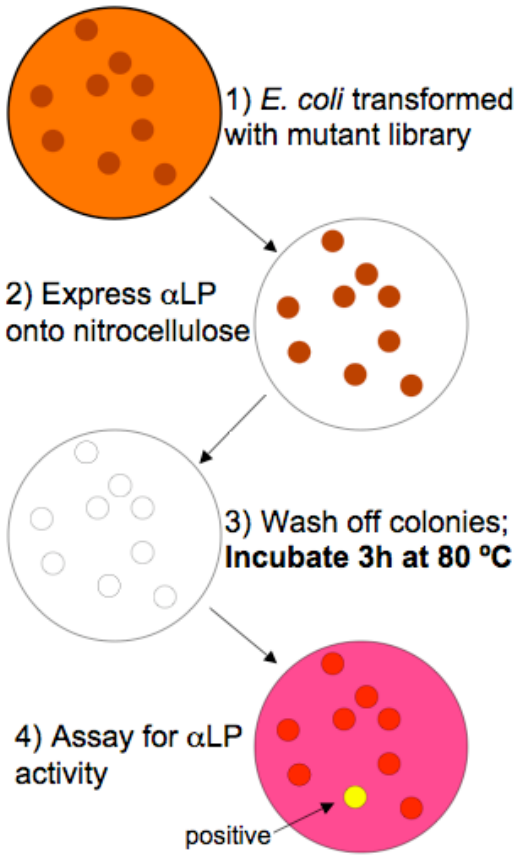


Figure 8.4: Schematic describing the screen for αLP mutants with slower unfolding.

To address this, we employed a genetic screen to identify mutants of αLP that would have larger unfolding barriers (Figure 8.4). Briefly, a mutant library of αLP plasmid (generated using the same mutagenesis protocol described above) was transformed into *E. coli* and grown on NZ-amine plates overnight to form colonies. These colonies were then lifted onto nitrocellulose filters and incubated in the presence of IPTG to induce production of αLP overnight. After the colonies were removed from the filter, the nitrocellulose was heated to 80 ºC for 3 hours to unfold the αLP protein attached to the filter. This treatment was sufficient to destroy a majority of the wt-αLP present, as measured in control reactions. However, if mutant αLP survived this treatment, perhaps through a larger unfolding barrier, its presence could be visualized using a chromogenic substrate assay.

Table 8.2: Mutants obtained in screen for slower unfolding αLP

Mutant*	Interesting Interactions	TFPA Cognate Residues‡
RH(48A)/S198P	Domain Interface	Q(48A)/**P198**
NK(120D)	Domain Bridge	**R(120D)**
SG(120H)	Domain Bridge	**G(120H)**
AT(48C)/A111P/T113I/SG(120H)	Domain Bridge	**S(48C)**/A111/T113/**G(120H)**
SR(120G)	Domain Bridge	N(120G)
SI(120G)/R178H	Domain Bridge/β-hairpin	N(120G)/T178
A176T	β-hairpin	**T176**
V218M	Mace Loop	None
V218L	Mace Loop	None
A131E/V218M	Mace Loop	A131/none

*chymotrypsin-homology numbering scheme
‡Bold indicates natural substitutions in TFPA recapitulated by mutations

This screen was employed on ~5000 clones, out of which 10 positive (αLP survivor) clones were obtained. The clones were sequenced to identify the mutations that resulted in surviving the heat treatment (Table 8.2). These mutations tend to cluster around the domain interface and especially in the domain bridge (Figure 8.5). This localization along the domain interface and domain bridge suggests that these regions are important in the unfolding process for αLP. Indeed, the unfolding TS has been proposed to involve a loss of interactions along the domain interface (Jaswal 2000)(Chapter 4) and the domain bridge (Chapter 6).

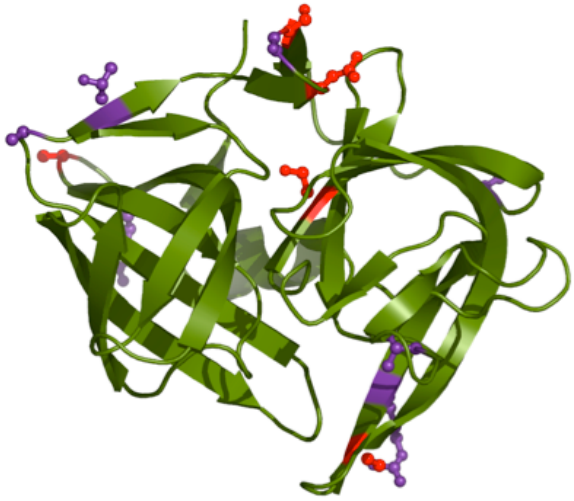


Figure 8.5: Location of mutants from high temperature screen.
Mutants are shown in ball and stick representation. The mutations that recapitulate recapitulate the sequence of TFPA are shown in red, while others are shown in violet. There is an enrichment of mutations occurring in the domain bridge, as many of these mutants were obtained multiple times.

Amazingly, many of the mutations obtained from random mutagenesis also recapitulate natural substitutions that are found in the thermophilic αLP homolog TFPA (Figure 8.5; Table 8.2). This observation would suggest that many of the residues that are crucial for TFPA's increased kinetic thermostability (Chapters 6 and 7) reside at these

mutated positions. This illustrates the power of an unbiased approach to identify interactions that candidates for increasing kinetic stability from the myriad differences between αLP and TFPA.

To test whether the heat-resistance qualities of the αLP mutants translated from the nitrocellulose filter to solution, they were overexpressed, purified and their unfolding rates at high temperature were measured. Surprisingly and disappointingly, the unfolding rates at 70 ºC were not significantly different from wt-αLP (Figure 8.6). Therefore, none of the mutants tested had the desired deceleration of the unfolding rate.

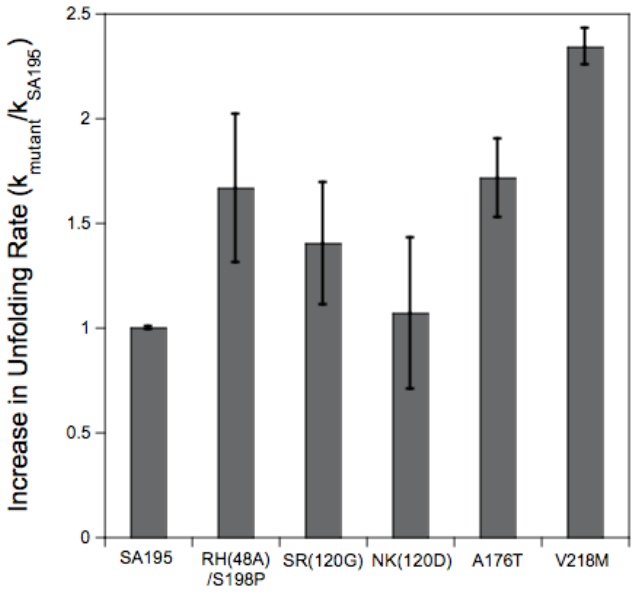


Figure 8.6: Effect on unfolding rate at 70 ºC of various mutants obtained from screen for higher unfolding barrier.
All mutants increased the unfolding rate rather than decreased it.

This begs the question, "Why did these mutants survive in the screen if their unfolding rates were essentially the same as wild-type?" One strong possibility is that the interaction with the nitrocellulose membrane caused major differences in the unfolding behavior and the mutations slowed this nitrocellulose-dependent unfolding but had no

effect on the solution unfolding rate. In support of this hypothesis, the length of time at 80 ºC required to destroy ~95% of the αLP activity on the nitrocellulose membrane (3 hours) is >100-fold slower than the time required to unfolding 95% of αLP in solution (~100 seconds), so there was certainly some sort of nitrocellulose-dependence to the unfolding. Alternatively, the heat served to wash the αLP off of the membrane and the mutations just increased the retention time of the protease on the membrane. I don't favor this hypothesis because the mutants survived the heat treatment even on dried membranes, whereas wild-type didn't. However, residual moisture could still serve to wash the membranes. Yet another hypothesis is that the mutants exhibit their effect not on the height of the unfolding barrier but on its cooperativity. In this case the heat caused a loss of cooperativity and αLP autolysis occurred on the membrane, resulting in less αLP being present to be active. The mutants, with their presumed greater cooperativity, prevented proteolysis from occurring and survived the heat treatment. This hypothesis is not favored because the protease molecules are immobilized on the membrane and would not be able to contact each other to conduct proteolysis. Finally, these mutants may have increased proteolytic activity such that the small amount of active protease remaining on the filter after heat treatment gives a sizable signal.

Regardless, the mechanism by which these αLP mutants attained their nitrocellulose-dependent heat resistance will require more study. However, the strategy of using directed evolution techniques to identify αLP mutants that have altered folding landscapes remains a valid and viable area of exploration. A solution-based screen for thermostability has been described which may provide a good blueprint for future screens(Arnold, Giver et al. 1999; Zhao and Arnold 1999).

Materials and Methods

Creation of the Mutant Libraries

Because the Derman protocol used hydroxylamine mutagenesis(Derman and Agard 2000), which only causes C ⇒ T transitions, many mutations were outside the range of this technique. To expand the mutagenic possibilities, we employed 2-stage mutagenic PCR strategy to create the mutant libraries, using the primers:

AlpSeqA (5'-CGACTTCGTCGCCCTCAGCG-3') and

AlpSeqZ (5'-GCTGTGACCGTCTCCGGGAGC-3').

The first round of mutageneis consisted of standard error-prone PCR using Taq polymerase in the presence of Mn^{2+} and skewed dNTP ratios, in this case 0.3mM Mn^{2+} and a 5-fold excess of dCTP and dTTP over dATP and dCTP (1mM vs. 0.2mM). This PCR was performed in 10mM Tris, pH 8.3, 50mM KCl, 7mM $MgCl_2$, and 0.01% (w/v) gelatin. This protocol causes a small number of mutagenic transversions as well as the much more common transitions, with a relatively low rate of deletions and insertions, but favors mutation at A and T bases. To make the mutagenesis as broad as possible, the PCR product was ethanol precipitated and then subjected to PCR using the commercially available Mutazyme I GeneMorph kit. This treatment causes relatively more transversion mutations and has a bias towards G and C, thus covering more possible sequence space. The mutagenesis was calibrated so that there were ~5 number of mutations per kilobase. This PCR product was ethanol precipitated and then digested with NcoI and XbaI and then ligated into the pALP12 vector similarly digested. The plasmid library was then

electroporated into *E. coli* strain AD1202 and plated onto NZ-amine agar (with carbenicillin) such that there were ~100 colonies per plate and grown at 37 ºC overnight.

A nitrocellulose filter (Schleicher and Schuell) was laid onto the colonies and fiducial markers were imprinted onto the master plate and filter so that the orientation of the filter can be matched to that of the plate. The filter is then lifted off the plate and then placed colony-side up onto an NZ-Carb plate that has been supplemented with 1mM IPTG.

Screening methods

For the screen for faster folding αLP using the Pro-3 variant, the plates containing the colony-loaded filters were placed at 25 ºC overnight. When using the Pro-3Ala variant, these plates were incubated at 18 ºC overnight. The filters were then lifted from the plates and colonies were washed off. αLP activity was then visualized using a previously described protocol(Mace and Agard 1995; Derman and Agard 2000).

For the screen for slower unfolding mutants, the plates containing the colony-loaded filters were incubated at 18 ºC overnight to induce αLP expression. The filters were lifted and the colonies washed off. Most of the moisture was removed from the filters and they were placed into hermetically sealed plastic bags. The bags were then submerged into 80 ºC water for 3 hours. Unfolding was quenched by submersion into ice cold water. The filters were removed from the plastic bags and then αLP activity visualized according to an established protocol(Mace and Agard 1995; Derman and Agard 2000).

Protein production and unfolding

Mutant protein was expressed and purified as previously described.(Mace, Wilk et al. 1995) Unfolding was performed as previously described (Sohl, Jaswal et al. 1998; Truhlar, Cunningham et al. 2004). All chemicals were purchased from Sigma Aldrich.

Chapter 9: Future Directions

Further Investigation of Strain in Folding

The strain from distortion of Phe228 appears to be playing a key role in the modulation of folding and unfolding behaviors in the large Pro region-dependent proteases. Further investigation of this phenomona is required to truly understand this process and how it can be modulated.

While the Repack3 mutant roughly recapitulates the free energy landscape of SGPB, it can only account for ~40% of the energetic differences between αLP and SGPB. This result is expected as the Repack3 mutant has only 3 residues mutated to their SGPB cognates, or only ~3% of the sequence variation between αLP and SGPB. We would expect that further mutation of Repack3 mutant to structurally resemble SGPB would increase its energetic resemblance as well.

Two residues that are prime candidates for further mutagenesis are Ile162 and Lys165. Both of these residues strongly covary according to pro region size((Fuhrmann, Kelch et al. 2004) and are either in direct contact with Phe228 (Ile162) or in the close vicinity (Lys165). Ile162 appears to be involved in the positioning of the nearby strand, which is extended away from Phe228 in SGPB. Conversion of this residue to Val, as found in SGPB, could reposition this strand and/or other core residues, causing better packing. This quadruple mutant has already been prepared (dubbed TIWLQIIV-αLP), but protein has not been expressed. Lys165 appears to be a key residue in positioning Thr181, as the long lysine sidechain is positioned across this residue and makes a conserved hydrogen bond with the carbonyl of Ala130. This interaction may hold

Thr181 in its tight contact with Phe228, which would generate the strain. Mutation of Lys165 (in the context of the Repack mutant) to leucine as found in SGPB may even further mimic the energetics of SGPB. These mutants could give further insight into the individual contributions of residues to strain generation and packing in the family of pro-dependent proteases.

Another way to investigate the role of packing in the generation of strain is to completely redesign the core of the protein using a computational design algorithm such as ROSETTA(Simons, Bonneau et al. 1999; Simons, Ruczinski et al. 1999). Some trials have been performed by Adam Thomason in conjunction with Tanja Kortemme while he was rotating in the lab. Adam used ROSETTA to pick residues that would optimally fill in the core of the C-domain. Each high-scoring sequence was modeled and then were minimized using the AMBER force-field(Ponder and Case 2003). Remarkably, the wild-type sequence was nearly always picked as the top scoring solution under many various redesign runs, regardless of how many or which residues were allowed to vary. In many cases, sequences that resembled the small Pro region proteases were amongst the highest scoring solutions as well, suggesting that the computational redesign strategy could provide interesting insight into the role of packing in kinetic stability. However, only rarely were these sequences were significantly different than those already made (such as Repack2 or Repack3.) For this reason, it may be worthwhile to redesign the αLP core using a flexible backbone strategy to more extensively explore conformational and sequence space for a sequence that optimally satisfies packing necessities without generating strain.

Although αLP is the best understood use of strain in modulating folding and/or unfolding, it may be a general feature of many proteins. Significant distortion of other Phe and Tyr residues has been observed in many other ultra-high resolution crystal structures ((Kang, Devedjiev et al. 2004) and Figure 3.1B). Therefore, distortion of covalent geometry in these proteins may also be used to functionally modulate protein energetics, possibly to generate kinetic stability as in αLP, but also for catalysis and/or allosteric transitions. Investigation of strain in other orthogonal systems may provide useful insight into the benefits and limitations of using strain to perform biological activity.

Finally, although we have been focused on the distortion of aromatic residues from planarity, deformation of covalent geometry in other types of bonds (such as aliphatic sidechains or the amide backbone) could be important and useful sources of strain as well. An analysis of the ultra-high resolution structures in the Protein Data Bank may provide a list of candidate proteins with which to study the potential role of strain in other types of bonds. This could provide interesting insights into the generality of strain as a tool for generating functional properties.

Further Investigation of the Structural Basis for Kinetic Stability

The structural analysis of NAPase and TFPA suggested that the Domain interface and the Domain Bridge play key roles in generating kinetic stability in the pro-dependent proteases. Subsequent mutagenesis experiments bear this hypothesis out and have led to a structural model for the unfolding Transition State for this entire class of proteins. While this model has provided key insights into the structural mechanisms of kinetic

stability, detailed knowledge of which specific interactions mediate this behavior is still lacking.

To understand the role of the domain interface in controlling unfolding in further detail, more mutations are necessary to probe specific interactions in dictating the unfolding process. While the Relocation mutants suggest that certain salt bridges (Arg141 with Glu32 and Arg103 with Glu229) are broken in the TS, these mutants did not show this unambiguously, as the observed effect on the pH dependence of unfolding may be a result of the newly introduced salt bridge which was placed distal to the domain interface. Single mutants are necessary that only remove the original salt bridges of αLP. These mutants currently are being generated by Pinar Erciyas and will yield important insight into what role these salt bridges play in the unfolding process. In addition, mutations of other residues along the domain interface will be important for understanding the extent of domain interface rupture in the TS.

The "cracked egg" model of αLP unfolding also predicts that there would be differential consequences of salt on the unfolding of αLP and NAPase. Since αLP has salt bridges that are broken in the TS but NAPase doesn't, then addition of salt should lower the αLP unfolding barrier due to Debye-Huckel screening of salt bridges. Indeed, when tested the αLP unfolding rate displays a steep dependence on salt whereas NAPase does not (Figure 5.3). Interestingly, the salt dependence at low pH was not as expected, which may be due to specific ion binding in the TS (Figure 5.4). Mutagenesis experiments could pinpoint the location of this putative ion binding behavior in a manner analogous to Psi Value analysis(Krantz and Sosnick 2001), which would give structural insight into the TS, as the ion binding seems to be TS-specific.

The cracked-egg model for unfolding also illustrates the importance of the domain bridge in dictating the unfolding process for these proteases. While the mutagenesis experiments and homolog comparisons outlined in Chapter 4, 5 and 6 clearly illustrate the importance of this interface, they don't clearly indicate which specific interactions are controlling unfolding. To examine this in more detail, further mutagenesis experiments are necessary with more focused attention paid to individual interactions. In addition, computational methods such as Molecular Dynamics unfolding simulations could be used to screen through conditions relatively quickly to identify important interactions that could be tested experimentally. In addition, simulations allows for interrogation of structural and energetic aspects that are inaccessible by experimental methods.

The correlation of the domain bridge interaction area with the unfolding rate indicates that the structural attributes of the domain bridge can be used in a predictive fashion to adjust the unfolding barrier. There are several other pro-dependent protease homologs with structures solved that could be used to predict the unfolding rate which could then be tested experimentally. In addition, this hypothesis could be further tested with addition chimeras, such as SGPB with the αLP domain bridge which would be expected to slow unfolding. Finally, an extended domain bridge computationally designed to optimally bind to the protein surface would be expected to significantly slow unfolding.

So far, we have models for most of the major states in the αLP folding landscape: the αLP:Pro Region complex, native αLP, and even the Transition State. However, the state which we have the least structural information about is the Intermediate (Int).

While pursuit of the αLP Int structure has been elusive for over 15 years, it may be possible to gain insight into this critical state by interrogating the structure of the NAPase and/or TFPA Intermediates. Study of the αLP Int has been refractory for many reasons, which may be alleviated when examined in either extremophile. First, the αLP Int has a tendency to aggregate at concentrations higher than 10μM, making it difficult or impossible to perform many experiments, such as NMR, which require concentrated solutions. Second, αLP Int aggregation also becomes an issue at temperatures higher than 10 ºC, which also seriously limits experimental investigation of Int structure. These aggregation problems may be alleviated with the NAPase and/or TFPA Int. Finally, folding of Int to native protease, although extremely slow, can complicate long experiments because the small amount of protease that folds during the time course can digest the remaining Int over a timescale of weeks. This is less likely to be an issue for NAPase and TFPA, as their folding rates are expected to be quite slow (see below).

Another poorly understood aspect of kinetic stability, and protein folding in general, is the mechanism of denaturant induced unfolding. Different models have been proposed to account for the chaotropic effects of denaturants, from disruption of water structure to binding directly to protein backbone and/or sidechains. To address this issue from a structural perspective, crystals of TFPA have been grown in two different ionic denaturants: sodium thiocyanate (NaSCN) and Guanidinium Hydrochloride (GdmHCl). This data provides a rare opportunity to examine the actual structural consequences of denaturant activity. Diffraction data for both conditions have been collected to high resolution (1.5Å in 0.4M NaSCN and 1.7Å in 0.2M GdmHCl), although neither dataset has been refined. Analysis of these datasets, with careful attention paid to the positioning

of denaturant molecules, could yield significant insight into the activity of denaturants. For example, identification and comparison of denaturant binding sites on the protein could illuminate whether denaturant molecules prefer to bind at certain moieties. Additionally, examination of water structure in the vicinity of the denaturant molecules could clarify the role of the water networks in protein stability and the effect that denaturants have on disrupting these networks.

Further Investigation of Extremophilic Behavior in Kinetically Stable Proteases

Extremophilic behavior is facilitated by kinetic stability, as exemplified by the structural and kinetic studies of both TFPA and NAPase (Chapter 4 and 6). While NAPase has been shown to exhibit an unfolding free energy barrier that is stable over a large pH range, it is not clear whether this quality comes at the price of some other feature. Likewise, the same is true for the remarkable kinetic thermostability of TFPA. One characteristic that has a good chance of being altered, especially in the case of NAPase, is the $\Delta C_p^{\ddagger}$, which describes the temperature dependence of the unfolding free energy barrier (N to TS). The $\Delta C_p^{\ddagger}$ reports on both the extent of non-polar surface area that becomes exposed and the amount of electrostatic interactions that are broken in the TS. Since it appears the NAPase TS minimizes the rupture of salt bridges, it would make sense that the $\Delta C_p^{\ddagger}$ would be altered from that of αLP. In addition, the $\Delta C_p^{\ddagger}$ of TFPA could be altered to slow unfolding at high temperature. Alterations of the ΔC_p, which measures the temperature dependence of ΔG_{N-U}, has been shown to be lowered in thermophilic proteins, thereby providing high temperature stability(Hollien and Marqusee 2002; Robic, Berger et al. 2002). An accurate determination of the $\Delta C_p^{\ddagger}$ for TFPA and/or

NAPase would provide significant insight into the nature of extremophilic behavior in both proteins as well as providing deep understanding into the unfolding mechanism.

In addition, an integral aspect of the TFPA and NAPase folding landscape hasn't yet been investigated: the folding reaction. How does the folding barrier of these extremophiles measure up against those of αLP and SGPB? It is expected that folding barriers would be increased since their unfolding barriers are larger (Figure 4.2; Figure 6.3). In addition, it would be of clear interest to understand the temperature and pH dependence of the folding rate to understand the response of the folding landscape to these perturbants. These studies would give useful information into the thermodynamic parameters and electrostatic components of the folding reaction.

To investigate the evolution and mechanisms of extremophilic behavior, directed evolution techniques could be used to identify mutations that confer resistance to certain perturbants. In the past, screens have been developed to identify heat-resistant proteases, as well as mutants that fold faster. Although there is more work that should go into development of screening methods (Chapter 8), this strategy could yield interesting insight into mechanisms of evolution of extremophilic behavior and kinetic stability for these proteases. Finally, a comparison using directed evolution for the same traits in a thermodynamically stable protein, such as trypsin, could reveal the relative adaptability of kinetic stability.

As an alternate and complementary approach to understanding evolution of extremophilic behcvior, studies of other extremophilic homologs should be undertaken as well. Currently, only one other known hyper-thermophilic (>90 ºC) homolog is known, which is from *Pyrococcus furiosus* (Pfu Protease A). I have cloned this protein out of the

P. furiosus genome and it has been subcloned into the pALP12 vector. Expression in *E.coli* D1210 was attempted but, without an assay for Pfu Protease activity, it is unknown if active protein was generated. Another homolog that will be interesting to investigate is from *Desulfitobacterium hafniense*, an organism that grows by chlororespiration on chlorinated phenols. The extracellular enzymes responsible for the dechlorination are highly reductive, suggesting that the environment that this protease has adapted to is quite reducing. Interestingly, this is the only Pro dependent protease that has no disulfides, as it contains no cysteine residues. Understanding the folding and unfolding of this protease could yield insight into the role of disulfides as well as redox conditions on kinetic stability. Regardless, more extremophilic homologs will be discovered, so attention should be paid to the sequence databank to find homologs that may display some interesting extremophilic behavior (psychrophiles, halophiles, alkaliphiles, etc.)

Finally, it may be possible to rationally engineer some novel extremophilic behavior into a mesophilic protein such αLP. This has already been accomplished for generating acidophilic and thermophilic behavior in αLP (Chapters 4 and 6, respectively). Novel forms of extremophilic character, such as alkaliphilia, could be generated through rational design based upon our structural knowledge of the αLP TS. This type of study would provide validation and refinement of current TS structural models, while also providing an interesting case of TS-based functional protein engineering.

Structural Basis for Extreme Cooperativity

αLP and SGPB have been shown to have extremely cooperative unfolding transitions(Jaswal, Sohl et al. 2002; Truhlar, Cunningham et al. 2004), but the mechanisms that underlie this behavior are still completely unknown. However, there is evidence that suggests that this ultra-cooperativity is not present in TFPA and/or NAPase. In gladiator assays at 37 ºC and pH 7, both of these proteases are degraded faster than αLP (Figure 4.4; Figure 7.5). This result was quite surprising because both NAPase and TFPA unfold significantly faster than αLP at higher temperatures (Figure 4.2; Figures 7.1 and 7.2) and, therefore, it was expected that these protein's slower unfolding would be extended to lower temperature as well. In fact, TFPA's extrapolated unfolding rate at 37 ºC is ~100 fold slower than αLP's (Figure 7.4). If the unfolding rates for TFPA and NAPase are indeed slower than αLP's at 37 ºC as expected, then it would suggest that their faster degradation is due to enhanced native state dynamics that render portions of the chain accessible to proteolysis. In support of this conclusion, inhibition of both TFPA and NAPase was required for crystallization (Chapters 4 and 6), suggesting that autolysis in the crystallization drop was preventing formation of ordered crystals.

So what is the structural basis for increased native state dynamics in TFPA and/or NAPase? The answer to this question could yield significant insight into the mechanism of αLP's and SGPB's ultra-cooperativity. By using mass spectrometry in conjunction with the gladiator assay, it may be possible to determine where in the proteins the initial cleavage reaction occurs. This may require careful consideration of conditions to ensure that only the initial cleavage event occurs and is not followed by subsequent cleavage reactions. Along these lines, the substrate specificity of TFPA and/or NAPase degradation could also suggest regions where the fluctuations are most prevalent. Also,

careful analysis of crystallographic B-factors may be able to determine regions of the protein that are inherently more dynamic. Finally, hydrogen-deuterium exchange, followed by either mass spectrometry or NMR, would be able to give very accurate determination of the dynamic properties of these two proteins.

Mechanisms of Pro-mediated Folding

Although a general model for the Pro-mediated folding of αLP has been formulated, the mechanistic details of this reaction are still largely unknown. It has been hypothesized that there is more direct contact between αLP and the Pro N-domain in the folding TS than in the native complex. This has led to the conclusion that the native αLP-Pro complex is strained such that the Pro N-domain interactions are less direct and energetically significant in the native state. To address this issue of strain, one could lengthen the linker between the Pro N- and C-domains so that this strain is relieved. Understanding the energetics of folding with this Pro variant could yield insight into the nature of the catalyzed TS. In addition, a crystal structure of the complex between this altered Pro with αLP could illuminate the αLP:Pro N-domain interactions that occur in the folding TS.

Also, it is unclear whether Pro assists in the distortion of Phe228 in the pro-catalyzed reaction. Does the removal of distortion increase the rate of catalyzed folding as well? This could be tested by measuring the catalyzed folding rate of the Repack mutant or T181G. If the pro-catalyzed rate is not effected by strain, then the conclusion would be that Pro accelerates folding simply through correct positioning of the αLP backbone and that the distortion is 'downhill' in the catalyzed folding landscape.

Finally, how do the TFPA and NAPase Pro Regions function under extreme conditions? The TFPA Pro Region must be able to function under extremely hot conditions, where the αLP Pro would be completely non-functional(Peters, Shiau et al. 1998). What is the mechanism of this stabilization? What are the energetic changes that have to be made for the TFPA Pro Region to catalyze folding at higher temperature? An in-depth analysis of TFPA's Pro mediated folding pathway could address these issues. Also, does the NAPase Pro Region catalyze folding under acidic conditions? It is not entirely clear if this is necessary and/or possible because the protease typically needs to be at neutral or basic pH to perform the cleavage reaction necessary to separate the protease and Pro-domain. An understanding of the catalyzed NAPase folding pathway under differing pH conditions could yield insight into the breadth of conditions conducive to NAPase folding as well as the electrostatic components of Pro-mediated folding.

General Mechanisms of Kinetic Stability

The members of the alpha-Lytic Protease subfamily (clan S2A according to the MEROPS protease classification system(Rawlings, Morton et al. 2006)) are not the only proteins that fold in a pro region dependent manner. Other cases of pro region dependent folding have been reported for the unrelated serine protease subtilisin BPN'(Eder, Rheinnecker et al. 1993; Bryan 2002). More detailed comparisons of the characteristics of αLP and subtilisin members could identify the benefits and limitations of different mechanisms of kinetic stability.

A particularly interesting case of pro dependent folding has been reported for penicillin amidase, a heterodimeric enzyme that catalyzes a similar reaction as a

protease(Ignatova, Mahsunah et al. 2003; Kasche, Galunsky et al. 2003; Ignatova, Wischnewski et al. 2005). It was found that the A and B subunits are synthesized as one polypeptide *in vivo* and that the intervening sequence is actually a pro region that accelerates folding by $>10^7$-fold(Ignatova, Wischnewski et al. 2005). The removal of the internal pro sequence requires at least 2 different cleavages that occur in an auto-catalytic manner. Understanding the mechanisms of pro dependent folding and kinetic stability of penicillin amidase will undoubtedly uncover unique and interesting biochemical phenomena.

Another case of pro dependent folding has been identified in lipases, extracellular enzymes that degrade lipids and are the most used enzymes for biotechnological purposes. Lipase requires an N-terminal pro region that catalyzes the folding of the enzyme(Frenken, Bos et al. 1993; Frenken, de Groot et al. 1993; El Khattabi, Ockhuijsen et al. 1999; El Khattabi, Van Gelder et al. 2000; Rosenau, Tommassen et al. 2004; Pauwels, Lustig et al. 2006; Rodriguez-Larrea, Minning et al. 2006). While high resolution structures exist for both lipase and the lipase:Pro complex(Pauwels, Lustig et al. 2006), there is no clear understanding of the mechanism of Pro dependent folding or the overall folding landscape. An understanding of these characteristics are important for understanding mechanisms of Pro dependent folding and kinetic stability in a non-protease, which will be of great interest to both the folding community and the biotechnological sector. For example, it is completely unclear how lipase sheds its foldase, as the foldase domain remains tightly bound as an inhibition complex upon completion of the folding reaction(Rosenau, Tommassen et al. 2004), similar to the Pro Region of αLP(Sohl, Shiau et al. 1997).

Structural Basis for Substrate Specificity

αLP has not only been a useful model for understanding folding mechanisms, but also enzyme catalysis(Bone, Shenvi et al. 1987; Fuhrmann, Daugherty et al. 2006) and substrate specificity(Bone, Frank et al. 1989; Bone, Silen et al. 1989; Bone, Fujishige et al. 1991; Bone, Sampson et al. 1991; Mace and Agard 1995; Mace, Wilk et al. 1995; Miller and Agard 1999; Ota and Agard 2001). αLP is specifically proteolyzes peptides with small hydrophobic residues at the S_1 position, which is believed to be due to a pair of methionine sidechains (from M190 and M213) that line the active site walls, forcing the inter-wall distance to be strictly enforced. Upon mutation of either residue to alanine, the enzyme loses specificity for small hydrophobic residues and will degrade peptides with nearly any residue at the S_1 position(Bone, Silen et al. 1989), indicating that the active site walls are no longer conformationally coupled(Bone, Silen et al. 1989; Bone, Fujishige et al. 1991; Miller and Agard 1999; Ota and Agard 2001). Since structural decoupling of the αLP active site walls removes all specificity, it would be expected that similar mutations in homologs would have the same result. However, the residues corresponding to M190 and M213 are much smaller (A190 and T213) in TFPA, NAPase, and SGPB, yet all of these enzymes have strong specificity for large hydrophobic and aromatic residues (data not shown). Because these proteases don't have broad specificity as in the αLP MA190 or MA213 variants, this suggests that there is still some conformation coupling in the active site walls of these proteins, although the source of this coupling is unknown. A combination of mutational, kinetic, crystallographic and

computational methods could yield insight into how the substrate specificity is determined in these enzymes.

References

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-30.

Arnold, F. H., L. Giver, et al. (1999). "Directed evolution of mesophilic enzymes into their thermophilic counterparts." Ann N Y Acad Sci **870**: 400-3.

Bai, Y., T. R. Sosnick, et al. (1995). "Protein folding intermediates: native-state hydrogen exchange." Science **269**(5221): 192-7.

Baker, D. and D. A. Agard (1994). "Kinetics versus thermodynamics in protein folding." Biochemistry **33**(24): 7505-9.

Baker, D., A. K. Shiau, et al. (1993). "The role of pro regions in protein folding." Curr Opin Cell Biol **5**(6): 966-70.

Baker, D., J. L. Sohl, et al. (1992). "A protein-folding reaction under kinetic control." Nature **356**(6366): 263-5.

Barrick, D. and R. L. Baldwin (1993). "Three-state analysis of sperm whale apomyoglobin folding." Biochemistry **32**(14): 3790-6.

Bender, M., Sturtevant, J.M. (1947). "The Heat of the Inactivation of Pepsin." J Am Chem Soc **69**(3): 607-612.

Bendtsen, J. D., H. Nielsen, et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." J Mol Biol **340**(4): 783-95.

Bone, R., D. Frank, et al. (1989). "Structural analysis of specificity: alpha-lytic protease complexes with analogues of reaction intermediates." Biochemistry **28**(19): 7600-9.

Bone, R., A. Fujishige, et al. (1991). "Structural basis for broad specificity in alpha-lytic protease mutants." Biochemistry **30**(43): 10388-98.

Bone, R., N. S. Sampson, et al. (1991). "Crystal structures of alpha-lytic protease complexes with irreversibly bound phosphonate esters." Biochemistry **30**(8): 2263-72.

Bone, R., A. B. Shenvi, et al. (1987). "Serine protease mechanism: structure of an inhibitory complex of alpha-lytic protease and a tightly bound peptide boronic acid." Biochemistry **26**(24): 7609-14.

Bone, R., J. L. Silen, et al. (1989). "Structural plasticity broadens the specificity of an engineered protease." Nature **339**(6221): 191-5.

Bonisch, H., C. L. Schmidt, et al. (2002). "The structure of the soluble domain of an archaeal Rieske iron-sulfur protein at 1.1 A resolution." J Mol Biol **319**(3): 791-805.

Botuyan, M. V., A. Toy-Palmer, et al. (1996). "NMR solution structure of Cu(I) rusticyanin from Thiobacillus ferrooxidans: structural basis for the extreme acid stability and redox potential." J Mol Biol **263**(5): 752-67.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Bryan, P. N. (2002). "Prodomains and protein folding catalysis." Chem Rev **102**(12): 4805-16.

Butler, J. S. and S. N. Loh (2005). "Kinetic partitioning during folding of the p53 DNA binding domain." J Mol Biol **350**(5): 906-18.

Butler, J. S. and S. N. Loh (2006). "Folding and misfolding mechanisms of the p53 DNA binding domain at physiological temperature." Protein Sci **15**(11): 2457-65.

Buzzell, A., Sturtevant, J.M. (1952). "The Heat of Denaturation of Pepsin." J Am Chem Soc **74**(8): 1983-1987.

Canet, D., M. Sunde, et al. (1999). "Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants." Biochemistry **38**(20): 6419-27.

Carter, P. and J. A. Wells (1988). "Dissecting the catalytic triad of a serine protease." Nature **332**(6164): 564-8.

Chen, H. M., J. L. You, et al. (1991). "Kinetic analysis of the acid and the alkaline unfolded states of staphylococcal nuclease." J Mol Biol **220**(3): 771-8.

Chenna, R., H. Sugawara, et al. (2003). "Multiple sequence alignment with the Clustal series of programs." Nucleic Acids Res **31**(13): 3497-500.

Chiti, F., N. Taddei, et al. (1999). "Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding." Nat Struct Biol **6**(11): 1005-9.

Clarke, J., E. Cota, et al. (1999). "Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway." Structure **7**(9): 1145-53.

Cooper, J. B., G. Khan, et al. (1990). "X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 A resolution." J Mol Biol **214**(1): 199-222.

Craik, C. S., S. Roczniak, et al. (1987). "The catalytic role of the active site aspartic acid in serine proteases." Science **237**(4817): 909-13.

Cunningham, E. L. and D. A. Agard (2003). "Interdependent folding of the N- and C-terminal domains defines the cooperative folding of alpha-lytic protease." Biochemistry **42**(45): 13212-9.

Cunningham, E. L. and D. A. Agard (2004). "Disabling the folding catalyst is the last critical step in alpha-lytic protease folding." Protein Sci **13**(2): 325-31.

Cunningham, E. L., T. Mau, et al. (2002). "The pro region N-terminal domain provides specific interactions required for catalysis of alpha-lytic protease folding." Biochemistry **41**(28): 8860-7.

D'Aquino, J. A., J. Gomez, et al. (1996). "The magnitude of the backbone conformational entropy change in protein folding." Proteins **25**(2): 143-56.

Daggett, V. (2006). "Protein folding-simulation." Chem Rev **106**(5): 1898-916.

Dalby, P. A., M. Oliveberg, et al. (1998). "Movement of the intermediate and rate determining transition state of barnase on the energy landscape with changing temperature." Biochemistry **37**(13): 4674-9.

de Los Rios, M. A. and K. W. Plaxco (2005). "Apparent Debye-Huckel electrostatic effects in the folding of a simple, single domain protein." Biochemistry **44**(4): 1243-50.

Debye, P., Huckel, E. (1923). "The interionic attraction theory of deviations from ideal behavior in solution." Phys. Z **24**: 185-206.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA, USA, DeLano Scientific.

Derman, A. I. and D. A. Agard (2000). "Two energetically disparate folding pathways of alpha-lytic protease share a single transition state." Nat Struct Biol **7**(5): 394-7.

Dimitriadis, G., A. Drysdale, et al. (2004). "Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump." Proc Natl Acad Sci U S A **101**(11): 3809-14.

Edelhoch, H. (1957). "The Denaturation of Pepsin. I. Macromolecular Changes." J Am Chem Soc **79**(23): 6100-6109.

Eder, J., M. Rheinnecker, et al. (1993). "Folding of subtilisin BPN': role of the pro-sequence." J Mol Biol **233**(2): 293-304.

El Khattabi, M., C. Ockhuijsen, et al. (1999). "Specificity of the lipase-specific foldases of gram-negative bacteria and the role of the membrane anchor." Mol Gen Genet **261**(4-5): 770-6.

El Khattabi, M., P. Van Gelder, et al. (2000). "Role of the lipase-specific foldase of Burkholderia glumae as a steric chaperone." J Biol Chem **275**(35): 26885-91.

Eriksson, A. E., W. A. Baase, et al. (1992). "Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect." Science **255**(5041): 178-83.

Favilla, R., A. Parisoli, et al. (1997). "Alkaline denaturation and partial refolding of pepsin investigated with DAPI as an extrinsic probe." Biophys Chem **67**(1-3): 75-83.

Fersht, A. (1999). Structure and Mechanism in Protein Science. New York, W.H. Freeman & Company.

Flanagan, M. A., B. Garcia-Moreno, et al. (1983). "Contributions of individual amino acid residues to the structural stability of cetacean myoglobins." Biochemistry **22**(25): 6027-37.

Fleming, P. J., N. C. Fitzkee, et al. (2005). "A novel method reveals that solvent water favors polyproline II over beta-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA)." Protein Sci **14**(1): 111-8.

Forrer, P., C. Chang, et al. (2004). "Kinetic stability and crystal structure of the viral capsid protein SHP." J Mol Biol **344**(1): 179-93.

Frenken, L. G., J. W. Bos, et al. (1993). "An accessory gene, lipB, required for the production of active Pseudomonas glumae lipase." Mol Microbiol **9**(3): 579-89.

Frenken, L. G., A. de Groot, et al. (1993). "Role of the lipB gene product in the folding of the secreted lipase of Pseudomonas glumae." Mol Microbiol **9**(3): 591-9.

Fuhrmann, C. N., M. D. Daugherty, et al. (2006). "Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis." J Am Chem Soc **128**(28): 9086-102.

Fuhrmann, C. N., B. A. Kelch, et al. (2004). "The 0.83 A resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain." J Mol Biol **338**(5): 999-1013.

Fuhrmann, C. N., Ota, N., Rader, S.D., Agard, D.A. (2003). Alpha-Lytic Protease. The Handbook of Proteolytic Enzymes. A. J. Barrett and N. D. Rawlings, Woessner, J.F.

Fujinaga, M., L. T. Delbaere, et al. (1985). "Refined structure of alpha-lytic protease at 1.7 A resolution. Analysis of hydrogen bonding and solvent structure." J Mol Biol **184**(3): 479-502.

Fuller, B. (1975). SYNERGETICS-Explorations in the Geometry of Thinking, Macmillan Publishing Co.

Fushinobu, S., K. Ito, et al. (1998). "Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: biased distribution of acidic residues and importance of Asp37 for catalysis at low pH." Protein Eng **11**(12): 1121-8.

Geierstanger, B., M. Jamin, et al. (1998). "Protonation behavior of histidine 24 and histidine 119 in forming the pH 4 folding intermediate of apomyoglobin." Biochemistry **37**(12): 4254-65.

Godoy-Ruiz, R., F. Ariza, et al. (2006). "Natural Selection for Kinetic Stability Is a Likely Origin of Correlations between Mutational Effects on Protein Energetics and Frequencies of Amino Acid Occurrences in Sequence Alignments." J Mol Biol **362**(5): 966-78.

Goto, Y. and S. Nishikiori (1991). "Role of electrostatic repulsion in the acidic molten globule of cytochrome c." J Mol Biol **222**(3): 679-86.

Gouet, P., E. Courcelle, et al. (1999). "ESPript: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-8.

Grantcharova, V. P., D. S. Riddle, et al. (2000). "Long-range order in the src SH3 folding transition state." Proc Natl Acad Sci U S A **97**(13): 7084-9.

Gunasekaran, K., S. J. Eyles, et al. (2001). "Keeping it in the family: folding studies of related proteins." Curr Opin Struct Biol **11**(1): 83-93.

Gusek, T. W. and J. E. Kinsella (1987). "Purification and characterization of the heat-stable serine proteinase from Thermomonospora fusca YX." Biochem J **246**(2): 511-7.

Hardy, F., G. Vriend, et al. (1993). "Stabilization of Bacillus stearothermophilus neutral protease by introduction of prolines." FEBS Lett **317**(1-2): 89-92.

Hedstrom, L., T. Y. Lin, et al. (1996). "Hydrophobic interactions control zymogen activation in the trypsin family of serine proteases." Biochemistry **35**(14): 4515-23.

Henikoff, S. and J. G. Henikoff (1992). "Amino acid substitution matrices from protein blocks." Proc Natl Acad Sci U S A **89**(22): 10915-9.

Higaki, J. N. and A. Light (1986). "Independent refolding of domains in the pancreatic serine proteinases." J Biol Chem **261**(23): 10606-9.

Hocker, B. (2005). "Directed evolution of (betaalpha)(8)-barrel enzymes." Biomol Eng **22**(1-3): 31-8.

Hofmeister, F. (1888). "Zur Lehre von der wirkung der salze, II." Arch. Exp. Pathol. Pharmakol. **24**: 247-260.

Hollien, J. and S. Marqusee (2002). "Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H." J Mol Biol **316**(2): 327-40.

Huang, K., W. Lu, et al. (1995). "Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with Streptomyces griseus proteinase B." Protein Sci **4**(10): 1985-97.

Hughson, F. M. and R. L. Baldwin (1989). "Use of site-directed mutagenesis to destabilize native apomyoglobin relative to folding intermediates." Biochemistry **28**(10): 4415-22.

Huyghues-Despointes, B. M., U. Langhorst, et al. (1999). "Hydrogen-exchange stabilities of RNase T1 and variants with buried and solvent-exposed Ala --> Gly mutations in the helix." Biochemistry **38**(50): 16481-90.

Huyghues-Despointes, B. M., J. M. Scholtz, et al. (1999). "Protein conformational stabilities can be determined from hydrogen exchange rates." Nat Struct Biol **6**(10): 910-2.

Ignatova, Z., A. Mahsunah, et al. (2003). "Improvement of posttranslational bottlenecks in the production of penicillin amidase in recombinant Escherichia coli strains." Appl Environ Microbiol **69**(2): 1237-45.

Ignatova, Z., F. Wischnewski, et al. (2005). "Pro-sequence and Ca2+-binding: implications for folding and maturation of Ntn-hydrolase penicillin amidase from E. coli." J Mol Biol **348**(4): 999-1014.

Ionescu, R. M. and M. R. Eftink (1997). "Global analysis of the acid-induced and urea-induced unfolding of staphylococcal nuclease and two of its variants." Biochemistry **36**(5): 1129-40.

Jacob, M. and F. X. Schmid (1999). "Protein folding as a diffusional process." Biochemistry **38**(42): 13773-9.

Jaenicke, R. (1991). "Protein stability and molecular adaptation to extreme conditions." Eur J Biochem **202**(3): 715-28.

James, M. N., A. R. Sielecki, et al. (1980). "Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 A resolution. A model for serine protease catalysis." J Mol Biol **144**(1): 43-88.

Jaswal, S. S. (2000). Thermodynamics, Kinetics and Landscapes in alpha-Lytic Protease: A Role for Pro Regions in Kinetic Stability. Department of Biochemistry and Biophysics. San Francisco, University of California, San Francisco**:** 155.

Jaswal, S. S., J. L. Sohl, et al. (2002). "Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability." Nature **415**(6869): 343-6.

Jaswal, S. S., S. M. Truhlar, et al. (2005). "Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases." J Mol Biol **347**(2): 355-66.

Johnson, S. M., R. L. Wiseman, et al. (2005). "Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses." Acc Chem Res **38**(12): 911-21.

Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). O: A Macromolecular Modeling Environment. Crystallographic and Modeling Methods in Molecular Design. C. E. Bugg, S., Springer-Verlag Press**:** 189-195.

Kang, B. S., Y. Devedjiev, et al. (2004). "The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography." J Mol Biol **338**(3): 483-93.

Karshikoff, A. and R. Ladenstein (2001). "Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads." Trends Biochem Sci **26**(9): 550-6.

Kasche, V., B. Galunsky, et al. (2003). "Fragments of pro-peptide activate mature penicillin amidase of Alcaligenes faecalis." Eur J Biochem **270**(23): 4721-8.

Kashiwagi, T., N. Kunishima, et al. (1997). "The novel acidophilic structure of the killer toxin from halotolerant yeast demonstrates remarkable folding similarity with a fungal killer toxin." Structure **5**(1): 81-94.

Kaushik, J. K., K. Ogasahara, et al. (2002). "The unusually slow relaxation kinetics of the folding-unfolding of pyrrolidone carboxyl peptidase from a hyperthermophile, Pyrococcus furiosus." J Mol Biol **316**(4): 991-1003.

Kitadokoro, K., H. Tsuzuki, et al. (1994). "Purification, characterization, primary structure, crystallization and preliminary crystallographic study of a serine proteinase from Streptomyces fradiae ATCC 14544." Eur J Biochem **220**(1): 55-61.

Kragelund, B. B., P. Hojrup, et al. (1996). "Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family." J Mol Biol **256**(1): 187-200.

Krantz, B. A. and T. R. Sosnick (2001). "Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil." Nat Struct Biol **8**(12): 1042-7.

Kristjansson, M. M. and J. E. Kinsella (1990). "Alkaline serine proteinase from Thermomonospora fusca YX. Stability to heat and denaturants." Biochem J **270**(1): 51-5.

Kristjansson, M. M. and J. E. Kinsella (1990). "Heat stable proteinase from Thermomonospora fusca. Characterization as a serine proteinase." Int J Pept Protein Res **36**(2): 201-7.

Kubelka, J., J. Hofrichter, et al. (2004). "The protein folding 'speed limit'." Curr Opin Struct Biol **14**(1): 76-88.

Kumar, S. and R. Nussinov (2001). "How do thermophilic proteins deal with heat?" Cell Mol Life Sci **58**(9): 1216-33.

Lao, G. and D. B. Wilson (1996). "Cloning, sequencing, and expression of a Thermomonospora fusca protease gene in Streptomyces lividans." Appl Environ Microbiol **62**(11): 4256-9.

Laskowski, R. A., M. W. Macarthur, et al. (1993). "Procheck - a Program to Check the Stereochemical Quality of Protein Structures." Journal of Applied Crystallography **26**: 283-291.

Lejbkowicz, F., R. Kudinsky, et al. (2005). "Identification of *Nocardiopsis dassonvillei* in a Blood Sample from a Child." American Journal of Infectious Diseases **1**(1): 1-4.

Li, R. and C. Woodward (1999). "The hydrogen exchange core and protein folding." Protein Sci **8**(8): 1571-90.

Light, A. and A. M. al-Obeidi (1991). "Further evidence for independent folding of domains in serine proteases." J Biol Chem **266**(12): 7694-8.

Lockless, S. W. and R. Ranganathan (1999). "Evolutionarily conserved pathways of energetic connectivity in protein families." Science **286**(5438): 295-9.

MacBeath, G., P. Kast, et al. (1998). "Exploring sequence constraints on an interhelical turn using in vivo selection for catalytic activity." Protein Sci **7**(2): 325-35.

Mace, J. E. and D. A. Agard (1995). "Kinetic and structural characterization of mutations of glycine 216 in alpha-lytic protease: a new target for engineering substrate specificity." J Mol Biol **254**(4): 720-36.

Mace, J. E., B. J. Wilk, et al. (1995). "Functional linkage between the active site of alpha-lytic protease and distant regions of structure: scanning alanine mutagenesis of a surface loop affects activity and substrate specificity." J Mol Biol **251**(1): 116-34.

Main, E. R., K. F. Fulton, et al. (1999). "Folding pathway of FKBP12 and characterisation of the transition state." J Mol Biol **291**(2): 429-44.

Martinez, J. C. and L. Serrano (1999). "The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved." Nat Struct Biol **6**(11): 1010-6.

Matouschek, A., J. T. Kellis, Jr., et al. (1989). "Mapping the transition state and pathway of protein folding by protein engineering." Nature **340**(6229): 122-6.

Matthews, B. W., H. Nicholson, et al. (1987). "Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding." Proc Natl Acad Sci U S A **84**(19): 6663-7.

Miller, D. W. and D. A. Agard (1999). "Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease." J Mol Biol **286**(1): 267-78.

Mitsuiki, S., M. Sakai, et al. (2002). "Purification and some properties of a keratinolytic enzyme from an alkaliphilic Nocardiopsis sp. TOA-1." Biosci Biotechnol Biochem **66**(1): 164-7.

Munoz, V., P. A. Thompson, et al. (1997). "Folding dynamics and mechanism of beta-hairpin formation." Nature **390**(6656): 196-9.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.

Muslin, E. H., S. E. Clark, et al. (2002). "The effect of proline insertions on the thermostability of a barley alpha-glucosidase." Protein Eng **15**(1): 29-33.

Myers, J. K., C. N. Pace, et al. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein Sci **4**(10): 2138-48.

Nakagawa, S. and H. Umeyama (1984). "Role of catalytic residues in the formation of a tetrahedral adduct in the acylation reaction of bovine beta-trypsin. A molecular orbital study." J Mol Biol **179**(1): 103-23.

Nakamura, S., T. Tanaka, et al. (1997). "Improving the thermostability of Bacillus stearothermophilus neutral protease by introducing proline into the active site helix." Protein Eng **10**(11): 1263-9.

Nienaber, V. L., K. Breddam, et al. (1993). "A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding." Biochemistry **32**(43): 11469-75.

Ohnishi, S. and K. Takano (2004). "Amyloid fibrils from the viewpoint of protein folding." Cell Mol Life Sci **61**(5): 511-24.

Ota, N. and D. A. Agard (2001). "Enzyme specificity under dynamic control II: Principal component analysis of alpha-lytic protease using global and local solvent boundary conditions." Protein Sci **10**(7): 1403-14.

Otwinowski, Z. M., W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. C. W. S. Carter, R.M. New York, Academic Press. **276:** 307-326.

Pace, C. N. (1986). "Determination and analysis of urea and guanidine hydrochloride denaturation curves." Methods Enzymol **131**: 266-80.

Pappenberger, G., C. Saudan, et al. (2000). "Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements." Proc Natl Acad Sci U S A **97**(1): 17-22.

Park, C. and S. Marqusee (2004). "Probing the high energy states in proteins by proteolysis." J Mol Biol **343**(5): 1467-76.

Park, C. and S. Marqusee (2005). "Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding." Nat Methods **2**(3): 207-12.

Pauwels, K., A. Lustig, et al. (2006). "Structure of a membrane-based steric chaperone in complex with its lipase substrate." Nat Struct Mol Biol **13**(4): 374-5.

Pedersen, J. S., D. E. Otzen, et al. (2002). "Directed evolution of barnase stability using proteolytic selection." J Mol Biol **323**(1): 115-23.

Perl, D., M. Jacob, et al. (2002). "Thermodynamics of a diffusional protein folding reaction." Biophys Chem **96**(2-3): 173-90.

Perl, D., C. Welker, et al. (1998). "Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins." Nat Struct Biol **5**(3): 229-35.

Peters, R. J., A. K. Shiau, et al. (1998). "Pro region C-terminus:protease active site interactions are critical in catalyzing the folding of alpha-lytic protease." Biochemistry **37**(35): 12058-67.

Plaxco, K. W., K. T. Simons, et al. (1998). "Contact order, transition state placement and the refolding rates of single domain proteins." J Mol Biol **277**(4): 985-94.

Ponder, J. W. and D. A. Case (2003). "Force fields for protein simulations." Adv Protein Chem **66**: 27-85.

Potterton, E., S. McNicholas, et al. (2002). "The CCP4 molecular-graphics project." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1955-7.

Potterton, L., S. McNicholas, et al. (2004). "Developments in the CCP4 molecular-graphics project." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2288-94.

Privalov, P. L. and G. I. Makhatadze (1990). "Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects." J Mol Biol **213**(2): 385-91.

Prodromou, C. and L. H. Pearl (1992). "Recursive PCR: a novel technique for total gene synthesis." Protein Eng **5**(8): 827-9.

Rader, S. D. and D. A. Agard (1997). "Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 A resolution." Protein Sci **6**(7): 1375-86.

Rawlings, N. D., F. R. Morton, et al. (2006). "MEROPS: the peptidase database." Nucleic Acids Res **34**(Database issue): D270-2.

Reiling, K. K., J. Krucinski, et al. (2003). "Structure of human pro-chymase: a model for the activating transition of granule-associated proteases." Biochemistry **42**(9): 2616-24.

Robic, S., J. M. Berger, et al. (2002). "Contributions of folding cores to the thermostabilities of two ribonucleases H." Protein Sci **11**(2): 381-9.

Rodriguez-Larrea, D., S. Minning, et al. (2006). "Role of solvation barriers in protein kinetic stability." J Mol Biol **360**(3): 715-24.

Rosenau, F., J. Tommassen, et al. (2004). "Lipase-specific foldases." Chembiochem **5**(2): 152-61.

Rupley, J. A. (1967). "Susceptibility to attack by proteolytic enzymes." Methods Enzymol **11**: 905-917.

Russell, R. J. and G. L. Taylor (1995). "Engineering thermostability: lessons from thermophilic proteins." Curr Opin Biotechnol **6**(4): 370-4.

Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol Biol Evol **4**(4): 406-25.

Sauter, N. K., T. Mau, et al. (1998). "Structure of alpha-lytic protease complexed with its pro region." Nat Struct Biol **5**(11): 945-50.

Scandurra, R., V. Consalvi, et al. (2000). "Protein stability in extremophilic archaea." Front Biosci **5**: D787-95.

Schafer, K., U. Magnusson, et al. (2004). "X-ray structures of the maltose-maltodextrin-binding protein of the thermoacidophilic bacterium Alicyclobacillus acidocaldarius provide insight into acid stability of proteins." J Mol Biol **335**(1): 261-74.

Schindler, T., P. L. Graumann, et al. (1999). "The family of cold shock proteins of Bacillus subtilis. Stability and dynamics in vitro and in vivo." J Biol Chem **274**(6): 3407-13.

Settembre, E. C., J. R. Chittuluru, et al. (2004). "Acidophilic adaptations in the structure of Acetobacter aceti N5-carboxyaminoimidazole ribonucleotide mutase (PurE)." Acta Crystallogr D Biol Crystallogr **60**(Pt 10): 1753-60.

Shindyalov, I. N. and P. E. Bourne (1998). "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path." Protein Eng **11**(9): 739-47.

Silen, J. L. and D. A. Agard (1989). "The alpha-lytic protease pro-region does not require a physical linkage to activate the protease domain in vivo." Nature **341**(6241): 462-4.

Silen, J. L., D. Frank, et al. (1989). "Analysis of prepro-alpha-lytic protease expression in Escherichia coli reveals that the pro region is required for activity." J Bacteriol **171**(3): 1320-5.

Simons, K. T., R. Bonneau, et al. (1999). "Ab initio protein structure prediction of CASP III targets using ROSETTA." Proteins **Suppl 3**: 171-6.

Simons, K. T., I. Ruczinski, et al. (1999). "Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins." Proteins **34**(1): 82-95.

Socolich, M., S. W. Lockless, et al. (2005). "Evolutionary information for specifying a protein fold." Nature **437**(7058): 512-8.

Sohl, J. L., S. S. Jaswal, et al. (1998). "Unfolded conformations of alpha-lytic protease are more stable than its native state." Nature **395**(6704): 817-9.

Sohl, J. L., A. K. Shiau, et al. (1997). "Inhibition of alpha-lytic protease by pro region C-terminal steric occlusion of the active site." Biochemistry **36**(13): 3894-902.

Sprang, S., T. Standing, et al. (1987). "The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis." Science **237**(4817): 905-9.

Stackhouse, T., J. J. Onuffer, et al. (1987). "The role of protein folding in the evolution of protein sequences." Cold Spring Harb Symp Quant Biol **52**: 537-44.

Suel, G. M., S. W. Lockless, et al. (2003). "Evolutionarily conserved networks of residues mediate allosteric communication in proteins." Nat Struct Biol **10**(1): 59-69.

Thirumalai, D., D. K. Klimov, et al. (2003). "Emerging ideas on the molecular basis of protein and peptide aggregation." Curr Opin Struct Biol **13**(2): 146-59.

Truhlar, S. M. (2004). Decoupled folding and functional landscapes allow alpha-Lytic Protease and Streptomyces griseus protease B to develop novel native-state

properties. Department of Biochemistry and Biophysics. San Francisco, University of California, San Francsico**:** 136.

Truhlar, S. M. and D. A. Agard (2005). "The folding landscape of an alpha-lytic protease variant reveals the role of a conserved beta-hairpin in the development of kinetic stability." Proteins **61**(1): 105-114.

Truhlar, S. M., E. L. Cunningham, et al. (2004). "The folding landscape of Streptomyces griseus protease B reveals the energetic costs and benefits associated with evolving kinetic stability." Protein Sci **13**(2): 381-90.

Vallee-Belisle, A., J. F. Turcotte, et al. (2004). "raf RBD and ubiquitin proteins share similar folds, folding rates and mechanisms despite having unrelated amino acid sequences." Biochemistry **43**(26): 8447-58.

Walter, R. L., S. E. Ealick, et al. (1996). "Multiple wavelength anomalous diffraction (MAD) crystal structure of rusticyanin: a highly oxidizing cupredoxin with extreme acid stability." J Mol Biol **263**(5): 730-51.

Watanabe, K., K. Chishiro, et al. (1991). "Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from Bacillus thermoglucosidasius KP1006." J Biol Chem **266**(36): 24287-94.

Watanabe, K., T. Masuda, et al. (1994). "Multiple proline substitutions cumulatively thermostabilize Bacillus cereus ATCC7064 oligo-1,6-glucosidase. Irrefragable proof supporting the proline rule." Eur J Biochem **226**(2): 277-83.

Winn, M. D., M. N. Isupov, et al. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Crystallogr D Biol Crystallogr **57**(Pt 1): 122-33.

Zhao, H. and F. H. Arnold (1999). "Directed evolution converts subtilisin E into a functional equivalent of thermitase." Protein Eng **12**(1): 47-53.

Zhu, G. P., C. Xu, et al. (1999). "Increasing the thermostability of D-xylose isomerase by introduction of a proline into the turn of a random coil." Protein Eng **12**(8): 635-8.

Appendix A: List of homologs in the alpha-Lytic Protease sub-family

Organism	Accession #	Pro Length	Name	Observations of Note
Acidovorax avenae	ABM31117	198	Protease II	plant pathogen
Acidovorax avenae	YP_969445	178	Protease I	plant pathogen
Arabidopsis thaliana	Q9LK70	176	Protease I	eukaryotic; large insertions
Clostridium perfringen	NP_563146	119	Protease I	
Clostridium phytofermentans	ZP_01352384	192	Protease I	no cysteines, bad signal sequence: intracellular?
Corynebacterium diphtheriae	NP_939116	132	Protease I	human pathogen
Corynebacterium efficiens	NP_737418	164	Protease I	Thermophile
Corynebacterium jeikeium	YP_251357	50	Protease I	human pathogen
Dehalococcoides sp.	CAI83121	0	Protease I	
Desulfitobacterium hafniense	YP_518248	193	Protease I	no cysteines; lives in reducing environment
Janibacter sp. HTCC2649	ZP_00994643	164	Protease I	
Lysobacter enzymogenes	ABI54136	166	aLP	Most studied member of class; pdb:1SSX
Metarhizium anisopliae	CAB44651	168	Protease I	Eukaryotic; Insect pathogen
Mycobacterium avium	ABK69095	0	Protease I	bird pathogen
Mycobacterium avium	NP_959517	0	Protease II	bird pathogen
Mycobacterium avium	AAS02723	0	Protease III	bird pathogen
Mycobacterium flavescens	ZP_01195411	0	Protease I	

Organism	Accession #	Pro Length	Name	Observations of Note
Mycobacterium flavescens	ZP_01195163	0	Protease II	
Mycobacterium leprae	NP_302490	0	Protease I	human pathogen
Mycobacterium smegmatis	ABK71514	0	Protease I	
Mycobacterium ulcerans	ABL06271	0	Protease I	human pathogen
Mycobacterium vanbaalenii	YP_953977	0	Protease I	
Myxococcus xanthus	YP_633609	163	Protease I	C-terminal Ricin domain
Nocardiopsis alba	AAO06113	168	NAPase	Acid-resistant; 2OAU.pdb
Phaeosphaeria nodorum	EAT90061	170	Protease I	Eukaryotic; plant pathogen
Pyrococcus sp. V222	AAY29758	150	Protease I	Hyperthermophile
Rarobacter faecitabidus	Q05308	167	RPI	C-terminal Ricin domain
Rhodococcus sp.	ABG96867	0	Protease I	pathogen
Rhodococcus sp.	YP_705420	180	Protease II	pathogen
Rhodococcus sp. RHA1	YP_705419	182	Protease III	pathogen; aLP-length Mace loop
Salinispora arenicola	EAX26256	180	Protease II	Marine organism; C-terminal Ricin domain
Salinispora arenicola	ZP_01650994	127	Protease I	Marine organism
Salinispora tropica	ZP_01430889	126	Protease I	Marine organism
Salinispora tropica	EAU22084	151	Protease II	Marine organism; C-terminal extension
Stenotrophomonas maltophilia	ZP_01642284	158	Protease I	Mace loop extension
Stigmatella aurantiaca	ZP_01461796	171	Lytic-protease	C-terminal Ricin domain
Stigmatella aurantiaca	EAU67417	171	Protease C	
Stigmatella aurantiaca	ZP_01464200	172	Protease I	C-terminal Ricin domain

Organism	Accession #	Pro Length	Name	Observations of Note
Stigmatella aurantiaca	ZP_01461852	163	Protease II	
Stigmatella aurantiaca	ZP_01466391	130	Protease III	
Streptomyces albogriseolus	BAA06163	73	Protease I	
Streptomyces albogriseolus	BAA21784	68	Protease II	
Streptomyces ambofaciens	CAJ88492	140	Protease I	extended Mace loop
Streptomyces ambofaciens	CAJ88582	176	Protease II	altered C-domain
Streptomyces ambofaciens	CAJ88510	146	Protease III	C-terminal extrension
Streptomyces avermitilis	NP_822175	167	Protease I	
Streptomyces avermitilis	NP_827728	75	Protease II	
Streptomyces avermitilis	NP_828673	140	Protease III	C-terminal extenstion
Streptomyces avermitilis	NP_827729	139	Protease IV	
Streptomyces coelicolor	NP_625129	174	Protease I	C-terminal Carbohydrate Binding Domain
Streptomyces coelicolor	NP_628830	165	Protease II	C-terminal Carbohydrate Binding Domain
Streptomyces coelicolor	NP_631438	169	Protease III	
Streptomyces coelicolor	NP_626019	77	Protease IV	
Streptomyces coelicolor	NP_625036	138	Protease V	C-terminal extenstion
Streptomyces coelicolor	NP_625055	145	Protease VI	C-terminal extenstion
Streptomyces coelicolor	NP_626014	137	Protease VII	
Streptomyces coelicolor	NP_630031	184	Protease IX	
Streptomyces fradiae	Q03424	137	SFaseI	pdb: 2SFA
Streptomyces griseus	P52320	167	SGPC	C-terminal Carbohydrate Binding Domain

Organism	Accession #	Pro Length	Name	Observations of Note
Streptomyces griseus	S37460	138	SGPE	Glutamate Specific; pdb: 1HPG
Streptomyces griseus	P00776	77	SGPA	pdb: 2SGA
Streptomyces griseus	P00777	77	SGPB	pdb: 4SGB
Streptomyces griseus	P52321	136	SGPD	
Streptomyces hygroscopicus	AAO61210	98	Protease I	
Streptomyces lividans	NC_003155	177	SALA	C-terminal Carbohydrate Binding Domain
Streptomyces lividans	CAD42808	174	SALB	C-terminal Carbohydrate Binding Domain
Streptomyces lividans	CAD42809	166	SALC	C-terminal Carbohydrate Binding Domain
Streptomyces lividans	BAA08785	122	SALD	
Streptomyces sp.	CAA52206	172	Alk. ProteaseI	
Streptomyces sp.	CAA52205	166	Protease I	
Streptomyces sp.	AAM96214	73	Protease II	
Thermobifida fusca	AAC23545	152	TFPA	Thermophile

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Brian Kuhl 3/29/07

Author Signature Date