Title
Thermodynamics, kinetics and landscapes in alpha-lytic protease

Author
Jaswal, Sheila,

Publication Date
2000

Peer reviewed|Thesis/dissertation

Thermodynamics, Kinetics and Landscapes in Alpha-lytic Protease:
A Role for Pro Regions and Kinetic Stability

by

Sheila S. Jaswal

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

[redacted]

Date University Librarian

Degree Conferred: .....

To: permissions@nature.com
From: Sheila Jaswal <baloo@itsa.ucsf.edu>
Subject: permission to reproduce for my thesis

Dear Ms Williams,
I would like to reproduce a paper I coauthored in Nature (Vol 395, pp. 817-819, 22 October 1998) as part of my thesis: Thermodynamics, Kinetics, and Landscapes: A Role for Pro Regions and Metastability in alpha-lytic Protease, to be submitted next month for my doctoral degree at the University of California, San Francisco. Please let me know what I need to do in order to obtain permission.
Thank you very much.
Sincerely,
Sheila Jaswal

From: "Williams, Marie" <M.Williams@nature.com>
To: "'baloo@itsa.ucsf.edu'" <baloo@itsa.ucsf.edu>
Subject: RE: Permission request
Date: Mon, 10 Jul 2000 10:52:10 +0100

10 July 2000

Dear Sheila Jaswal,

Nature, hereby grants permission for the reproduction of a figure from Nature 395:817-819 (1998)', providing that the authors agree and the following credit line is used:

Reprinted by permission from Nature (reference citation) copyright (year) Macmillan Magazines Ltd.

This is for print versions only.

Yours sincerely,

Marie Williams
Nature Permissions

I dedicate this thesis to

my father and mother,

Sitaram and Alice (the original Drs. Jaswal),

and my partner Meg Robertson.

Acknowledgements

After seven years of working on and thinking about barriers and landscapes, it is inevitable that I would come to characterize my progress through life and graduate school in those terms...

I would like to thank my parents, Sitaram and Alice for providing me with my first landscape, which was a thermodynamically stable one due to their presence. They have given me so much love and support from the beginning, encouraging me to try many different "conformations" and helping make every opportunity possible accessible to me. They continue to offer their advice and love, and support me in whichever landscape I find myself. My brothers, Paul, Rog, and Vikram played important parts in that initial landscape as well, mostly in making up part of the barrier to maturation. However, since then, they have provided love and support, for which I am grateful. Vikram has been especially wonderful to have nearby in the South Bay, and provided excellent younger brother support and encouragement. I also need to acknowledge some great positive role models in my Math and Science teachers in the Lincoln Public Schools of math and science, who were mostly women, and got me turned on to and confident with science at an early age.

Following the childhood landscape in Lincoln, my next landscape at Mills College in Oakland, California was full of challenges and barriers, but was overall hugely stabilizing. I want to thank my professors in the Biology and Chemistry Departments for excellent teaching and endless time spent answering my questions and encouraging my curiosity. Dr. John Vollmer relentlessly pushed me to take part in every possible event

encouraging women in science. Dr. Keeports made Physics and especially Physical Chemistry unbelievably exciting and deserves credit for my current fascination with thermodynamics. I learned a lot from Sandra Banks, who was an extremely rigorous teacher. Dr. Brabson was great at encouraging questioning and wondering, and practical advice about pursuing graduate school. Dr. Margarethe Kulke, Dr. Karen Swearingen, Dr. John Harris, and Dr. Spiller helped me learn and discover enthusiasm for biology, despite my tendency to become overwhelmed by anything larger than a molecule.

With help from them, I got into summer research internship programs, where I learned about doing experiments. Dr. James Van Etten at the University of Nebraska provided me with two great summer undergraduate research experiences learning molecular biology. Dr. Simon Carding and Dr. Laila McVay at the University of Pennsylvania gave me independent molecular biology projects that let me experience guiding my own research. A year at the Max Planck Institute for Immunobiology in Freiburg working on transcription in Dr. Karl-Heinz Klempnauer's lab prepared me for the realities of research in graduate school (somewhat...) and convinced me to join a thesis lab where no molecular biology would be necessary!

Once I began graduate school at UCSF, I was thrust into a strongly denaturing environment, where the days of a thermodynamically stable landscape were over. Although I progressed from a very unfolded 1st year state into a more stable state after joining the Agard lab, that was in fact analogous to the αLP intermediate state: high in entropy, fairly inactive (at times) and facing a huge barrier to the native state (graduation). Despite the fact that graduate students don't enter graduate school with an

attached pro region, I was lucky to find many, many people who helped lower this barrier for me.

In my first rotation, Peter Walter was very supportive and provided very helpful advice for my fellowship applications. Carol Gross, my third rotation advisor, was hugely important in helping me survive graduate school in the early years. Her support was invaluable in helping me believe in myself and my scientific passion, despite lacking confidence. She spent hours calming me from panic and helping me rehearse every overhead for one of my early talks to turn it into a successful experience. She has stayed very involved in my progress on my orals and thesis committees, and has provided many crucial pep talks.

Susan Marqusee from the Molecular Biology and Biochemistry Department at Berkeley was very generous in serving on my thesis committee and making the trip to UCSF for each meeting. It has been great to have her expert folding insight and input, as well as getting me hooked up with her students who were thinking about similar issues. In particular I had several stimulating discussions with Aaron Chamberlain and Manuel Llinas about some Hydrogen Exchange experiments that unfortunately didn't materialize in this landscape.

Ken Dill really helped me to think rigorously about thermodynamics since the beginning, and has helped me to go beyond what most people try to get out of unfolding data. In the last few months he has been extremely generous with his time: spending time in several multiple hour meetings on denaturant binding and transition state hydration, thinking about and emailing almost daily about my analysis. He has been very

enthusiastic and gracious about being coerced into collaborating with us to come up with the theory to go with our experimental results.

On the practical front, Joyce Ramponi has shown infinite patience and willingness to help me with everything from finding phone numbers, to fixing the copier, to making 26 copies of my fellowship five minutes before the Fed Ex deadline and hand delivering it to the office. Sue Adams is always amazingly on top of everything and always had what I needed or knew the answer to every question- and with a cheerful, welcoming smile. Julie Ransom mothered me even though I wasn't officially a Biophysics student.

My involvement with Women in Life Sciences provided a lot of support and validation and perspective in the early to middle years. I had great experiences running the mentoring program and hosting Women Leaders in Science. It was wonderful to be involved with having a positive impact on the UCSF community as well as making friends with a bunch of really excellent women: Liz Haswell, Julia Charles, Tania Gonzalez, Sharon Stranford, Mary O'Riordan, Julia Owens, Tannishtha Reya, Julie Strong, Amy Kistler, Brooke Yool, Donna Hendrix, Lisa Uyechi, Renee Williard, Teresa Gamble.

I spent a really rewarding year as a scientist co-leader of the Presidio Middle School Girls' Science Club, run through the UCSF Science Education Partnership's Triad Program. That was another great outlet outside of lab, with the tangible reward of sharing science with girls. I learned a huge amount from that experience and the workshops run by the SEP staff (Liesl Chatman, Kimberly Tanner, Erin Strauss, Helen Doyle, Tracy Stevens).

I couldn't have made it though graduate school without a bunch of "snack buddies" who kept me fed and were good listeners during our coffee/snack breaks. Katherine Lemon was a great scientific sounding board during my first year and was always up for snack! Christina Hull and Jennifer Harris each were especially valuable in that role during the dark period of the intermediate years. Nira Pollock was always cheerful and willing to laugh at my jokes. Katrin Stade and Nina Boedekker provided support, humor and a chance to improve my German during coffee breaks. Tannisththa Reya never failed to make me laugh. Linda Brinen has provided valuable perspective and advice during this last transition year. I've really appreciated Marc Lenburg's zany take on life, and his help in maintaining perspective, as well as his patient Mac advice.

Dara Spatz Friedman was a housemate during several of the intermediate years. She has been a hugely generous and supportive friend, who provided critical listening and support helping me keep it together during very dark times.

Amy Kistler and I met during the interviews for Berkeley and UCSF, and immediately recognized kindred spirits. We've been roommates for all seven years, and have evolved from "Team Graduate Student", virtually inseparable outside of our respective labs during our first year, to dear companions. Her flexibility, insight, support and above all humor have helped keep me sane during these years.

I feel incredibly blessed to have spent my graduate years in the Agard Lab. From the beginning it has felt like a family, made up of a wonderful group of great scientists and fun, nice people. Everyone has been completely supportive of me, and my recent frequent requests (demands...) for help.

Julie Sohl is my "Sohl" sister. We're both from Nebraska, but kindred spirits for so many more reasons. I would not be finishing my thesis on αLP now without her enthusiasm for αLP folding and her patience with teaching me in the first few years since my rotation where I got all the great results that got me jazzed enough to stick it out through the next three years of stagnation! She has been amazingly supportive and provided great fun as a lab mate : torturing Michael with angry women music, starting the tradition of "poke"doc. She has remained a wonderful friend outside of lab- going skiing and hiking (stay away from fire trails with Julie!) and is supportive of all I do.

For Alan Derman, I am the little sister he's glad he never had... We started in lab at the same time, we're leaving at the same time, we've been through the hills and valleys together. He has put up with endless amounts of torture- poking, insults, subjection to Janet Jackson and "The Millenium Hip Hop Collection ". He could always be persuaded to get a snack or meal. He was always willing to think about my experimental problems with me, and to help work out spec or HPLC problems. He has been a great friend, and a critical component of the lab being a fun place to be, despite his occasional wet blanket outbursts.

Erin Cunningham is the little sister I always wished I had. She is my partner in crime: in torturing Alan, in being tortured by the CD and Spec, continuing the tradition of strong women in the Agard lab, listening to good music and making lab fun. We've also had many adventures outside of lab including skiing, hiking, video making. I want to thank her for coming to my rescue with thesis formatting during the last hectic days.

Stephanie Truhlar just joined the lab, but we shared several adventures en route to and at the "Protein Folding in the Cell" meeting in Vermont this summer. I am incredibly

grateful to her for dedicating days of her time dealing with all aspects of thesis formatting and proofreading. I would not have stayed sane and made it through the final painful push without her generous and cheerful help.

Cynthia Fuhrman is the nicest person I know, despite my best efforts to corrupt her to meanness. She is always cheerful and upbeat, and I owe her many thanks for helping me with random Mac questions, helping me learn Powerpoint, and scanning in and helping me with photos for my slides.

Along with Alan, Ted Mau and Andy Shiau made up the Night Crew. They were a 10th floor institution, working into the early hours. I spent a lot of time with them in the early years. Andy was a fount of wisdom and knowledge about all things scientific, especially practical experimental tips. He was good for lots of good troubleshooting about experiments, and some crucial senior grad student advice and perspective and listening during some very dark times. He also shared my appreciation for dance mixes, especially late at night! Ted was very good at thinking about experimental problems and being supportive. I managed to corrupt him out of some of his seriousness to come dancing at the Box, go on hikes, and got him hooked on Dance Mixes.

Reuben Peters was always friendly and cheerful despite sleeping through many of my group meetings. He and I did crystal structures of mutants early on, and he helped me a lot dealing with refinement and figure making. Stephen Rader worked with me during my rotation to express, purify, crystallize and solve the structure of SA195. I appreciated his vast scientific curiosity about all sorts of things, and his help and perspective. Marco Tonelli spent a huge amount of time last summer helping me do HX and 1D NMR timepoints, coming in early and staying late, but always remaining cheerful

and patient. Nick Sauter put up with a lot of abuse, but took it very good naturedly. He was also a big fan of the Courtyard Caffe Cream Cheese Brownie, and spent a lot of time helping me with figures.

Barry Wilk (aka Barry J. Blige) made the SA195 construct, and helped with SA195 expression and purification and trials for ^{15}N expression. Jim Mace gave lots of technical advice, especially on purification. Jill Nephew was my first rotation student, who asked lots of hard questions and was a lot of fun to work with. We did the ANS binding experiments together. Yoko Shibata was always willing to help solve any manner of problem, and to go get coffee. She provided good support and perspective, and is the master of organizing lab social events, along with Prabha Dias. Seth Harris has provided crucial humor during my months of painful thesis writing. Nobu Ota and David Miller were computational gurus, who helped me think about thermodynamics and helped write programs for me.

Michelle Moritz was a great senior presence to have in the lab in addition to Alan. She is cheerful, funny, and friendly, however, she is the mastermind of the dark side of the lab (using EM and working on centrosome and chromosome structure). Her secret plan is to take over the lab, as more and more people joining the lab are seduced to the dark side. The problem is, the members of the dark side are some of nicest people you'll ever meet.

Prabha Dias is always good for coffee, and has very practical suggestions for science and non science issues. I am indebted to her for lots of help with digital pictures, scanning, and photoshopping. Hector Aldaz and Matt Trammell are always interested in my experiments and hearing about what's going on scientifically, despite their own lack

of interest in αLP, the most interesting protein in the world. Peter Koenig shared his Talking Heads CD with me, and occasionally let me practice my German. Michael Braunfeld put up with sharing a lab with me for 6 years.

One of the two major pro regions in my graduate landscape has been my thesis advisor, David Agard. He has been an absolutely fantastic mentor, going above and beyond the norm for graduate advisors. He has spent much of my graduate years writing programs for me, meeting with me at the drop of a hat, driving to Montara to meet with Ken and me to talk about denaturant binding during his vacation week, spending hours indulging my obsession with gory thermodynamics. I have learned so much from him scientifically. From the beginning he treated me like an equal, and valued my ideas even before I did. I admire his fearless curiosity- that there must be a way to figure things out, and his depth of understanding about anything scientific.

Most of all, I owe him tremendous respect and gratitude for valuing me as a whole person first, and a scientist second. He has always made it clear that my emotional health was without questions most important. He has been amazing in acknowledging and understanding how hard my time in graduate school has been, and has done everything possible to help me on all levels. This has really allowed me to maintain perspective and balance, while actually enhancing my passion for science.

Finally, I want to acknowledge the love and affection and tremendous support of my partner, Meg Robertson. I couldn't have done this without her being there for me for the past five and a half years. Now that we have gotten married, I have the best pro region I could possibly have, for life!

Thermodynamics, Kinetics and Landscapes in α-lytic Protease:

A Role for Pro Regions and Kinetic Stability

Sheila S. Jaswal

Laboratory of Dr. David A. Agard

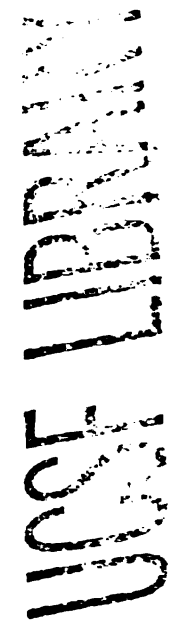
The folding of the extracellular serine protease, α–lytic protease (αLP) reveals a novel mechanism for stability that appears to lead to a longer functional lifetime for the protease. We demonstrate that for αLP, stability is based not on thermodynamics, but on kinetics: both an intermediate and the completely unfolded molecule are more stable than the native state. This has required the co-evolution of a pro region to facilitate folding by stabilizing both the folding transition state and the native state. The pro region is then proteolytically degraded, leaving the active αLP trapped in a metastable conformation by a large barrier to unfolding. The result has been the optimization of native state properties independent of their consequences on thermodynamic stability.

Despite its position at the global free energy minimum of the αLP folding landscape, the intermediate state is confirmed as a molten globule through exposure of hydrophobic surface area, and shown to aggregate, possibly in an ordered manner, at high temperatures. It is 50% folded, based on burial of surface area, and is stabilized entropically over the native state.

Further exploration of αLP's conformational landscape leads to a thermodynamic understanding of kinetic stability and a physical model for the high energy transition state. Additionally, based on the complete determination of the unfolding kinetics of

αLP, we demonstrate that current models for analysis of unfolding by denaturant are inadequate for describing the temperature and guanidine chloride dependent changes in the unfolding data.

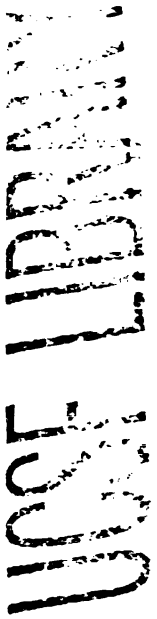
The transition state is highly native-like, with more than 80% surface area buried. Folding to the transition state leads to an entirely entropic penalty (at 10 °C), while enthalpy is favorable. Unfolding to the transition state results in a large $\Delta Cp_{N\text{-}TS}$, indicating significant exposure of hydrophobic surface area. This leads to an overall unfavorable entropy change, due to ordering solvent around that newly exposed surface in the transition state, which outweighs the gain in configurational entropy (at 10 °C). However, the transition state appears to have similar enthalpic stabilization to the native state. Therefore we propose a model for the transition state consisting of the two domains primarily natively folded, but separated at their interface, creating a solvent-exposed crevice, leading to an entropic but not enthalpic basis for αLP's kinetic stability.



Finally, we demonstrate that αLP has an enhanced ability to resist proteolysis and autoproteolysis under highly degradatory conditions. Therefore we propose that αLP's metastability may be a consequence of pressure to evolve properties of the native state, including a large, highly cooperative barrier to unfolding, and extreme rigidity, that reduce susceptibility to proteolytic degradation. Kinetic stability as a means to longevity is likely to be a mechanism conserved among the majority of extracellular bacterial pro-proteases, and may emerge as a general strategy for extracellular eukaryotic proteases subject to harsh conditions as well.

Table of Contents

List of Tables



List of Figures

Introduction

Background

Virtually all extracellular bacterial proteases are synthesized as precursor molecules with pro regions. In every case where the function of the pro region has been investigated, it has been found to be necessary for folding and secretion (Baker et al., 1993). One of the most striking and best studied examples of pro-mediated folding is the bacterial enzyme, α-lytic protease. α-Lytic protease (αLP) is a 198 residue serine protease secreted by the gram negative soil bacterium *Lysobacter enzymogenes* to degrade other soil microorganisms.

The overall three-dimensional fold of αLP clearly places it in the same family as the mammalian digestive serine proteases chymotrypsin, trypsin and elastase, despite only moderate sequence homology (Brayer et al., 1979). In contrast to these mammalian homologues, whose small N-terminal zymogen peptides simply prevent premature activation, αLP is synthesized with a large 166 residue N-terminal pro region (Pro) (Figure 1) that is required for proper folding of its mature protease domain (Silen, 1989).

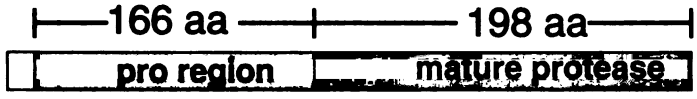


Figure 1. αLP precursor.

In vivo, co-expression of αLP and Pro, either in *cis* as the natural precursor molecule, or in *trans* as two separate polypeptide chains, results in the secretion of active αLP,

whereas expression of αLP alone leads to accumulation of the protease in the outer membrane due to apparent misfolding (Silen and Agard, 1989) (Figure 2).

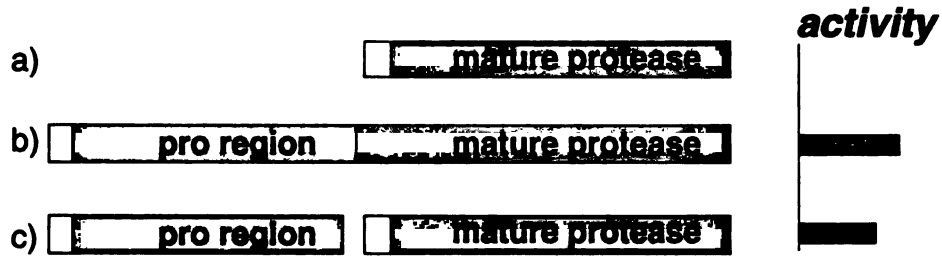


Figure 2. Requirement of pro region *in vivo*. a) If just the mature protease is expressed, no active enzyme is produced. However, if the pro region is provided, either b) in *cis*, as in the natural precursor enzyme, or c) in *trans* as a separate polypeptide, high levels of active enzyme are recovered.

In vitro refolding studies demonstrated that this requirement of the pro region is for proper folding of the protease. When chemically denatured, then diluted into physiological conditions, αLP refolds to an inactive intermediate, instead of the native conformation (Baker et al., 1992b). This intermediate persists for months, only converting to the active enzyme on the time-scale of millennia (Sohl et al., 1998). With the addition of the pro region, folding to the native conformation occurs in seconds (Baker et al., 1992b).

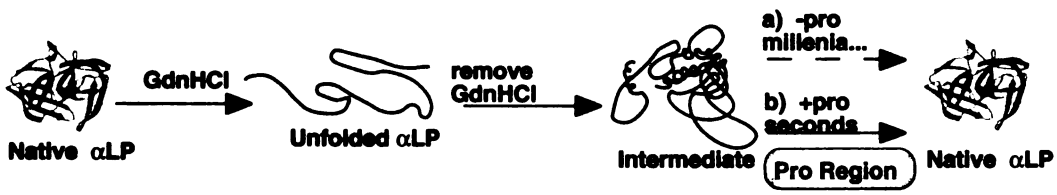
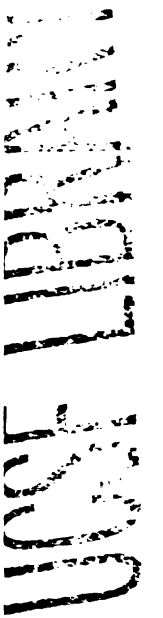


Figure 3. Requirement of pro region *in vitro*. Refolding in the absence of pro leads to an intermediate a) which only folds to native αLP in millennia. b) Addition of pro to intermediate leads to rapid refolding to native αLP.

The fact that the intermediate and native states exist under the same conditions without inter-converting in the absence of the pro region indicates that they are separated by a large kinetic barrier. To surmount this high barrier to folding, αLP has co-evolved the pro region, which can assist the folding of αLP when supplied in *cis* or *trans*. Emerging structural and energetic details of pro-mediated folding may define a theme for the folding of a wide range of homologous extracellular proteases that also contain pro regions.



Energetics of Pro-mediated Folding

Addition of Pro to I results in rapid folding to the N state (0.037 sec^{-1}) and recovery of functional protease (Fig. 4; (Sohl et al., 1998)). Pro acts as a foldase, facilitating αLP folding by binding tightly to the folding transition state of the protease, lowering the barrier by 18.2 kcal/mol. In this manner, Pro serves as a potent catalyst, increasing the rate of αLP folding by $3x10^9$. In addition, Pro is the tightest binding inhibitor known for the native protease ($K_i = 3x10^{-10}$ M; (Peters et al., 1998), (Baker et al., 1992a)), making Pro a single turnover catalyst. The product of the folding reaction is not active αLP, but the inhibitory complex. Release of active αLP requires that the Pro region be removed by proteolysis, which occurs naturally via αLP or another co-secreted protease.

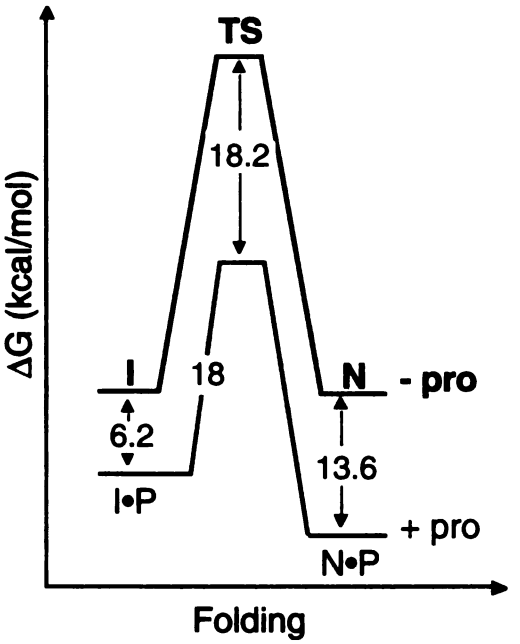


Figure 4. Free energy diagram of αLP folding. In the absence of its pro region (P) at 4 °C, unfolded αLP (U) spontaneously folds to a molten globule-like intermediate (I), which proceeds at an extremely slow rate to the native state (N) through a high energy folding transition state (TS). The addition of pro region provides a catalyzed folding pathway (denoted by blue lines) that lowers the high folding barrier and results in a thermodynamically stable inhibition complex (N•P).

Structures of Pro and Pro•αLP complex

Recently determined crystal structures of Pro and the Pro•N complex illuminate Pro•αLP interactions (Sauter et al., 1998). Alone, Pro adopts a novel C-shaped α/β fold, consisting of an N-terminal helix, two compact globular domains (N-domain, C-domain) connected by a nearly rigid hinge region and a C-terminal tail (Fig. 5a). Each globular domain contributes a three-stranded β-sheet to the concave surface of the molecule and at least one α-helix which packs against these β-sheets to form the convex surface. The N-terminal helix appears highly flexible, changing orientations in different crystal environments. Two of the three Pro molecules in the crystallographic asymmetric unit show different conformations for the N-terminal helix while the third molecule reveals

the helix to be disordered. Similarly, the C-terminal tail is unseen in the Pro structure, apparently disordered in unbound Pro.

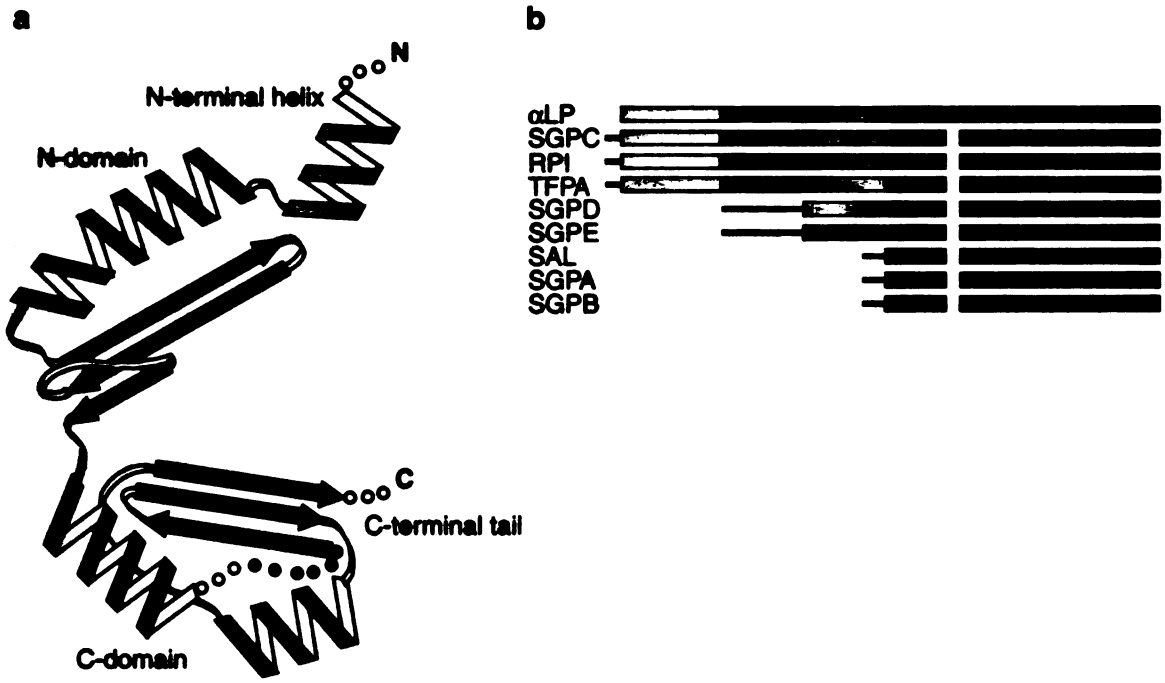
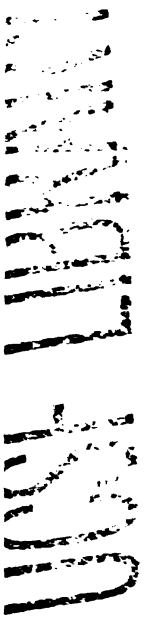


Fig. 5 a) Topology of Pro as described in the text. A disordered loop in the Pro C-domain is shown in red. b) Schematic of primary sequence alignments of pro regions from 9 bacterial serine proteases. Alignments were determined using the αLP Pro structure as a guide. Regions of sequence homology correspond to specific secondary structures in the Pro structure, with the Pro C-terminal domain being the most conserved region. N-terminal sequences lacking homology are depicted by thin black lines. αLP, *L. enzymogenes* α-lytic protease (Silen et al., 1988); SGPC, *S. griseus* protease C (Sidhu et al., 1994); RPI, *R. faecitabitus* protease I (Shimoi et al., 1992); SGPD, *S. griseus* protease D (Sidhu et al., 1995); SGPE, *S. griseus* protease E (Sidhu et al., 1993); TFPA, *T. fusca* serine protease (Lao and Wilson, 1996); SAL, *S. lividans* protease (Binnie et al., 1996); SGPA, *S. griseus* protease A (Henderson et al., 1987); SGPB, *S. griseus* protease B (Henderson et al., 1987).

Sequence comparisons with homologous pro-proteases suggest that the Pro structure may be a common pro region fold. Primary sequence alignments of Pro and eight related pro regions (Fig. 5b) indicate that despite a wide range of pro region sizes, these homologous

pro regions share common secondary structure elements, the most conserved region being that of the Pro C-terminal domain. These pro regions appear compatible with the Pro structure and presumably exhibit similar mechanisms of foldase activity.

In the case of αLP, Pro•N complex formation does not significantly alter the Pro structure (Fig. 6a). This is surprising since the pro region by itself has quite limited stability (T_m = 25 °C, 2.3 kcal/mol; (Anderson et al., 1999)), while Pro•N complex is greatly stabilized (13.6 kcal/mol; (Sohl et al., 1998)). However, the only notable differences are seen in the structuring of the C-terminal tail and the positioning of the flexible N-terminal helix upon protease binding. As expected for a tight binding inhibitory complex, the Pro•N complex structure buries a very large surface (> 4,000 Å^2) in its intermolecular interface.



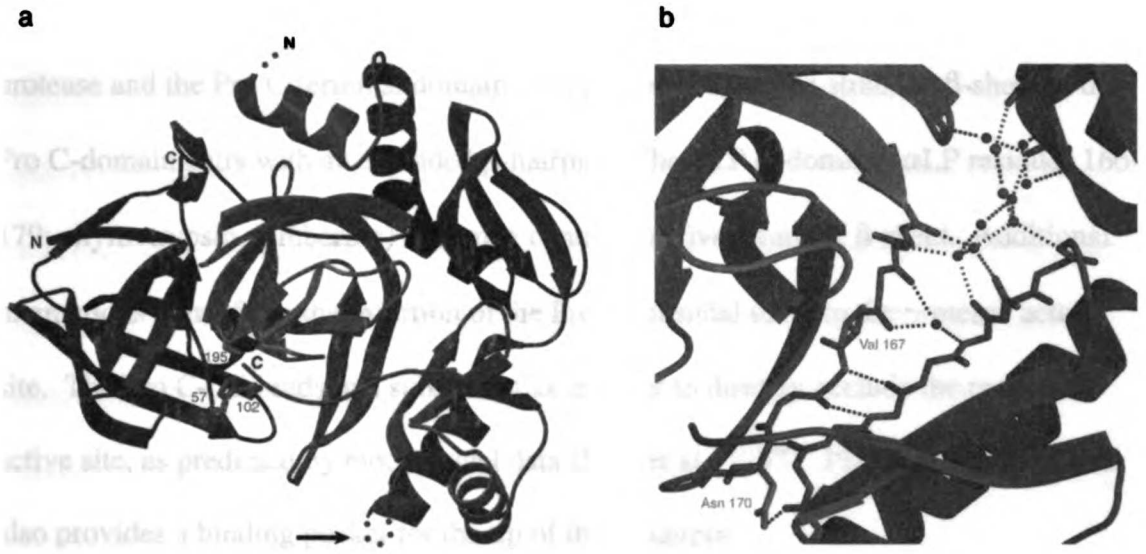
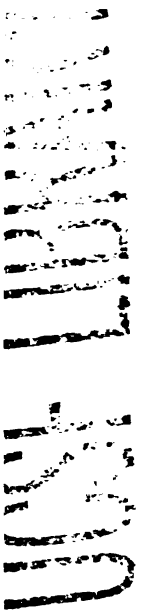


Fig. 6 a) Ribbon diagram of the Pro•N complex structure. The αLP N- and C-domains are colored magenta and blue respectively, with the side chains of the catalytic triad shown in red (His57, Asp102 and Ser195; chymotrypsin numbering). Illustrated in green, bound Pro inserts its C-terminal tail into the protease active site. A disordered loop in the Pro C-terminal domain, indicated by an arrow, presents a likely secondary protease cleavage site, leading to the release of active αLP from the inhibitory complex. b) Detail of the hydrated Pro•N interface. A gap between Pro (green) and the αLP C-domain (blue) is filled by ordered water molecules which are shown as red spheres. Some of these waters mediate hydrogen bonds (dashed orange lines) between the αLP β-hairpin and the Pro three stranded β-sheet which form the shared five-stranded b-sheet of the Pro•N interface. Residues in the αLP β-hairpin that affect formation of the initial Pro•I Michaelis complex (Ile167 and Asn170) are displayed in yellow. Figures modified from figures 2b and 3b (Sauter et al., 1998).

The most striking feature of the complex structure is the fact that Pro binds almost exclusively to the αLP C-domain, effectively surrounding the αLP C-terminal β-barrel. This observation raises the distinct possibility that it is the αLP C-domain that cannot fold properly and is therefore the focused substrate of Pro foldase activity. In support of this, recent mutagenesis studies indicate that the structuring of the protease C-domain is an integral part of the high folding barrier. Screens of libraries of chemically mutagenized αLP reveal that mutations that lower the folding transition state (as much as 3 kcal/mol) all map to the C-domain of the protease (Derman and Agard, 2000). The

most extensive and complementary interactions in the Pro•N interface occur between the protease and the Pro C-terminal domain. In particular, the three-stranded β-sheet in the Pro C-domain pairs with an extended β-hairpin in the αLP C-domain (αLP residues 166-179; chymotrypsin numbering) to form a continuous five-stranded β-sheet. Additional interactions come from the insertion of the Pro C-terminal tail into the protease active site. The Pro C-tail binds in a substrate-like manner to directly occlude the protease active site, as predicted by biochemical data (Sohl et al., 1997). Placement of the C-tail also provides a binding pocket for the tip of the β-hairpin.



αLP folding barrier and Pro foldase mechanism

The integration of prominent features of the complex structure with mutagenesis studies on both Pro and αLP provides significant insights into the origin of the folding barrier and the mechanism of Pro-catalyzed folding. Because Pro acts as a folding catalyst, it is possible to use modified Michaelis-Menten kinetics to extract functional information about the folding reaction (Peters et al., 1998). This analysis provides information on the formation of the Pro•I Michaelis complex (K_m) and the stabilization of the folding transition state (k_{cat}). In addition, the stability of the Pro•N complex can be assessed by measuring the inhibition of peptide substrate hydrolysis by Pro (K_i).

Mutations within the αLP β-hairpin loop alter both K_m and k_{cat} (Peters et al., 1998). The K_m effects reveal that formation of the shared β-sheet must occur in the first step of Pro catalyzed folding, while the k_{cat} effects indicate that this extended sheet continues to play a role during folding catalysis. Unlike these hairpin mutations, removing residues from the Pro C-tail (Peters et al., 1998) does not affect initial binding

to the αLP I state (K_m) and only marginally affects αLP N state binding (K_i), despite the Pro C-tail's high complementarity to the αLP binding pocket. In marked contrast, these same Pro C-tail truncations drastically reduce the folding rate (k_{cat}), profoundly hindering the ability of Pro to stabilize the folding transition state (TS). Deletion of the last three residues from the Pro C-tail decreases k_{cat} approximately 300-fold and removal of an additional fourth residue decreases folding by at least a factor of 10^7. The Pro C-tail therefore plays a direct role in Pro foldase activity, preferentially stabilizing the folding TS over the I and N states. Preliminary data indicate that the Pro N-domain also contributes to the catalytic activity of Pro. Mutations in Pro at the protease-Pro N-domain interface affect TS stabilization (E.L.C., P. Chien, D.A.A., unpublished).

The folding transition state and the native state are likely to share many structural features. The extremely tight binding of the rigid Pro region to both the native state (13.6 kcal/mol) and the transition state (18.2 kcal/mol) suggests that at least the αLP C-domain must be similarly structured in both states. However, they cannot be identical. Although the Pro C-terminal tail makes ideal substrate-like interactions with αLP, deletions only minimally affect the stability of Pro•N, while causing profound effects on the folding transition state (Peters et al., 1998). This suggests that the native Pro•N complex must be "strained" such that the total binding energy possible for the Pro C-tail is not realized in the Pro•N complex. By contrast, the intrinsic binding energy of the Pro C-tail does seem to be fully realized when complexed to the folding transition state, since it is stabilized by an additional 5 kcal/mol compared to Pro•N (Fig. 4).

Observations based on the Pro•N complex structure (Sauter et al., 1998), suggest this strain may be the result of poor complementarity in regions of the Pro•N interface

which could be improved to yield the observed additional stabilization in the Pro•TS complex. Most notably, there is a significant gap in the interface where the protease meets the junction of the two Pro domains. This gap contains eight ordered solvent molecules, three of which act to mediate hydrogen bonds between the αLP β-hairpin and Pro β-strand. Such highly solvated interfaces have been previously observed where two surfaces interact in two different conformational states. These "adapter" waters are seen in protein-DNA complexes (Schirmer and Evans, 1990), (Gewirth and Sigler, 1995) where waters populate the interface in non-specific complexes, yet are excluded in the specific complex. Similarly, waters are often used to adapt quaternary changes in allosteric enzymes (Royer et al., 1996)), with fewer waters in the higher affinity state due to improved surface complementarity. The Pro•αLP TS may be similarly stabilized by excluding the bound waters, thereby reducing the entropic cost of ordering the waters and increasing the direct Pro•αLP interface. As the structure of free αLP is nearly identical to that of αLP complexed with Pro, it is probable that strong αLP N state interactions prevent optimization of the Pro•αLP interface predicted in the TS complex. Destabilizing αLP mutations may disrupt these interactions enough to distort the Pro-αLP complex towards more TS-like binding.

αLP folding model

This structural and mutagenesis data can be synthesized into a model of Pro catalyzed folding of αLP (Fig. 7; (Sauter et al., 1998)). In this folding scheme, the N- and C-domains of the expanded molten globule folding intermediate are separated, and the β-hairpin is exposed to solvent. Pre-folded Pro initiates protease folding by binding

to the hairpin, forming a continuous five-stranded β-sheet. Efficient folding requires the Pro C-tail to then bind to the nascent active site, positioning the hairpin, and thereby assisting the structuring of the αLP C-domain. Finally, the αLP N-domain docks and folds against the C-domain to complete both the catalytic triad and the packing of the N state core.

Studies of the intact Pro-αLP precursor (Anderson et al., 1999) support the proposed two step folding model. Precursor refolding experiments show biphasic kinetics, with an initial fast rate equal to the rate of pro folding alone, followed by a slower rate for pro-mediated folding of αLP.

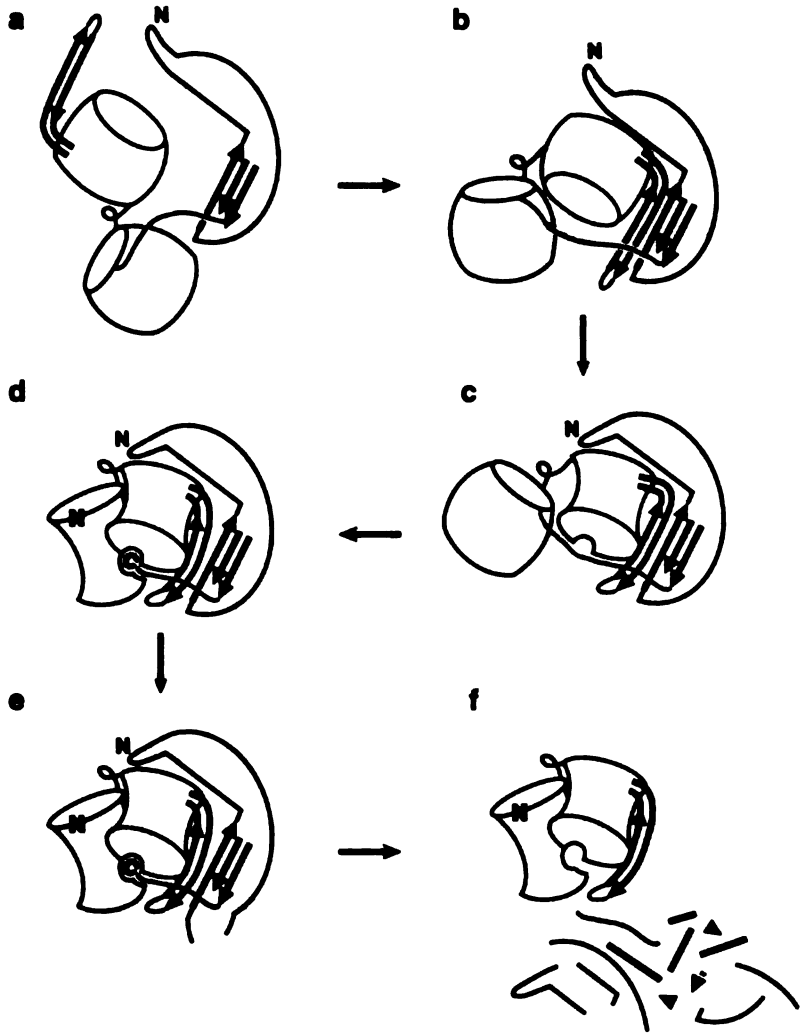


Fig. 7. Proposed model of Pro catalyzed folding of αLP. a) The pro domain of the Pro-αLP precursor folds, while the protease N- and C-domains remain separated and expanded. b) The three stranded b-sheet of the Pro C-domain pairs with the solvent exposed b-hairpin of the αLP C-domain, forming a continuous five stranded b-sheet. c) Substrate-like binding of the Pro-αLP junction to the nascent active site positions the b-hairpin and leads to the structuring of the αLP C-domain. d) The αLP N-domain folds upon docking with the αLP C-domain to complete the protease active site which can then process the Pro-αLP junction. The Pro C-terminal tail remains bound to the active site in this inhibitory complex while the new αLP N-terminus repositions to its native conformation. e) Intermolecular cleavage of secondary cleavage sites by αLP or other exogenous proteases leads to the f) eventual degradation of Pro and release of active, mature αLP. Color scheme as in figure 5. Figure modified from figure 4 (Sauter et al., 1998).

In *cis* folding, formation of the active site allows the protease domain of the precursor to process the Pro-αLP junction, producing the two distinct polypeptides chains

of the Pro•N complex. After cleavage, the Pro C-tail remains bound to the active site while the newly formed protease N-terminus repositions to its native conformation 24 Å away. Although the Pro-αLP precursor and Pro•N complex show similarities in secondary and tertiary structure, the marginal stability of the precursor (2.2 kcal/mol; (Anderson et al., 1999)) compared to the complex (10.6 kcal/mol) suggests that the rearrangement of the N-terminus is critical to αLP N-state stabilization. In addition to the primary intramolecular cleavage site, αLP also recognizes intermolecular cleavage sites within Pro, eventually leading to Pro degradation and release of active, mature protease. The disordered loop within the Pro C-domain (Figs. 6a, 7e), presents a likely target for the requisite secondary cleavage event. This secondary cleavage site is sensitive to many other proteases besides αLP. In fact, there may be a functional synergism between the multiple proteases secreted simultaneously by the host, *Lysobacter*, in cleaving each other's pro regions.

In the proposed folding scenario, Pro must bind to and correctly position the β-hairpin. The likely importance of this β-hairpin to the αLP folding barrier is reflected in its selective conservation among related proteases within the chymotrypsin superfamily. The β-hairpin, a common structural motif found in all 13 bacterial homologues synthesized with pro regions, is noticeably absent in other related bacterial, viral and mammalian proteases that do not require pro regions for proper folding. Furthermore, the Pro β-strand that pairs with the hairpin loop is also highly conserved in homologous pro regions, suggesting that the hairpin and its interaction with the Pro C-domain are important in structuring the protease. This is consistent with the fact that smaller related pro regions show sequence homology only to the Pro C-domain, thereby maintaining the

core structure necessary for binding the β-hairpin of the protease (Fig. 5b). The positioning of the hairpin and subsequent structuring of the αLP C-domain may be a general mechanism for pro region mediated folding of β-structures.

In contrast, subtilisin, a pro-protease evolutionarily unrelated to αLP, seems to employ a different method of pro catalyzed folding. The subtilisin pro domain stabilizes a pair of α-helices in the protease instead of a β-hairpin (Gallagher et al., 1995). Although αLP and subtilisin have convergently evolved pro-dependent folding, they differ in both their mature protease structures and the method by which their respective pro regions achieve their active protease conformations.

Parts of the preceding sections were used in a review written by Erin Cunningham and myself: "Kinetic Stability as a Mechanism for Protease Longevity", published as a Colloqium paper in *PNAS* 96: 11008-11014.

Physical and evolutionary origins of the folding barrier

Although the above studies provide insight into how the pro region is able to effectively guide folding to the native state, the fact remains that the pro region is immediately destroyed following folding. Therefore, for the functional lifetime of the active protease, αLP's conformational landscape is dominated by the kinetic barrier separating two very stable but very different conformations: the native and intermediate states. In order to understand the thermodynamic basis for the inability of αLP to fold on its own, as well as the evolutionary significance of pro dependent folding, it is necessary to thoroughly study the components of the folding landscape in the absence of the pro region and the barriers between them. Not only will these results shed light on the

general folding landscapes of pro-proteases, but features of αLP's kinetic barrier may provide insight into other proteins with multiple stable conformations of biological importance.

This thesis focuses on the folding of αLP in the absence of the pro region, in order to understand the thermodynamic and evolutionary basis for pro-dependent folding in αLP. The kinetic barrier between the native and intermediate states prevents them from being in equilibrium. What is the actual relative stability and breakdown of energetic contributions to the stability of the native and intermediate states? What is the nature of the long-lived but inactive intermediate? What is the progression of folding in terms of burial of surface, entropic and enthalpic contributions to the components and the kinetic barrier? Finally, can the thermodynamics and kinetics of the landscape along with the dependence on pro region be rationalized in terms of the biological function of αLP?

The first chapter describes work with Julie Sohl leading to the striking finding that the native state is less thermodynamically stable than the intermediate and unfolded states; therefore it is preserved solely by kinetic stability. This not only distinguishes αLP from its mammalian homologues but provides compelling support for the possibility of metastable native conformations in general. The nature of the unusually stabilized intermediate state is further revealed as definitively molten globule through ANS binding in Chapter 2.

Chapter 3 takes advantage of the wide range of conditions used to determine thermodynamic parameters from unfolding kinetics to demonstrate the inadequacies of common models for analysis of denaturant unfolding data in accurately describing temperature and denaturant dependent effects. Chapter 4 combines the analysis of the

accurately extrapolated unfolding data with temperature, with Julie's previous thermodynamic results to provide a nearly complete picture of the folding pathway in terms of burial of surface and enthalpic and entropic contributions at each step. Chapter 5 details the discovery of aggregation behavior of the intermediate under conditions of low denaturant and high temperature, inviting speculation on the possibility of ordered aggregation leading to fibril formation by αLP. Chapter 6 tests the hypothesis that the effects of these unusual thermodynamics on the native state of the protease have actually led to an increased ability to resist proteolysis and autoproteolysis. These results greatly illuminate the physical origins of kinetic stability as well as the rationale for the coevolution of kinetic stability and a pro region as a means to avoid self destruction.

Chapter 1. Unfolded Conformations of α–Lytic Protease are more stable than its native state.

Preface

After the establishment of a system for *in vitro* refolding by David Baker, it was revealed that the denatured protease could be refolded to the active enzyme in the presence of the pro region, but without the pro region, refolded instead to an inactive intermediate (Baker et al., 1992b). Although it displays typical characteristics of a marginally stable molten globule intermediate, this intermediate persists for months without converting to the active enzyme at a level capable of detection by the standard paranitroanilide hydrolysis assay. It remains capable of being refolded with the pro region for up to two months after its formation. The fact that the intermediate and native states exist under the same conditions without rapidly inter-converting indicated that a large kinetic barrier separated the two conformations. Based on the sensitivity of the paranitronanilide assay the refolding barrier was calculated to be at least 26 kcal/mol (Baker et al., 1992b). This obviously ruled out the presence of an equilibrium between the native and intermediate states, and revealed that one of these two states must be kinetically trapped.

Julie Sohl and I set out to determine which state was more stable, and to characterize thoroughly the complete energetics of the folding reaction. Julie determined the stability of the intermediate relative to the unfolded state using equilibrium urea denaturation. However, because the native and intermediate states are not at equilibrium, we measured the rates of refolding the intermediate and unfolding the native state to access the equilibrium free energy between them.

To validate the use of first order kinetics to analyze the refolding kinetics, we had to demonstrate the intermediate was monomeric, since it had been shown by gel filtration to be 25% expanded relative to the native state. Using equilibrium analytical ultracentrifugation, I showed that the intermediate is monomeric between 1 – 10 μM.

Julie developed a highly sensitive assay for native protease activity using a thiobenzyl ester substrate, which could detect the tiny fraction of intermediate: 1 in 10^6, which did refold in the absence of pro region over several days. Using this assay, she and I carried out refolding experiments to determine the rates of refolding at 4 and 25 °C.

In order to measure unfolding without the complications of autolysis at the high denaturant concentrations required for unfolding the native enzyme, I expressed and purified an inactive variant of αLP which has the active site serine replaced with an alanine (SA195). With Stephen Rader, I crystallized this and solved the structure by molecular replacement, and found that the SA195 structure is virtually identical to the wildtype structure (RMSD Cα's 0.173Å) (Rader, thesis). I followed the denaturation of SA195 at a series of guanidine hydrochloride concentrations at 4 and 25 °C to extrapolate the rate of unfolding in the absence of denaturant

As described in the following paper, reprinted with permission from Nature 395:817-819 © 1998, Macmillan Magazines Ltd., the results of these experiments led to the discovery that the native state of αLP is not at the global energy minimum, but is in fact less stable than both the intermediate and the completely unfolded states.

Unfolded conformations of α-lytic protease are more stable than its native state

Julie L. Sohl*†‡, Sheila S. Jaswal‡§ & David A. Agard*§

* *Graduate Group in Biophysics and the Howard Hughes Medical Institute,*
§ *Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448, USA*
‡ *These authors contributed equally to this work*
† *Present address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA.*

α-Lytic protease (αLP), an extracellular bacterial protease, is synthesized with a large amino-terminal pro-region that is essential for its folding *in vivo* and *in vitro*[1,2]. In the absence of the pro-region, the protease folds to an inactive, partially folded state, designated 'I'. The pro-region catalyses protease folding by directly stabilizing the folding transition state (>26 kcal mol^{-1}) which separates the native state 'N' from I[1,3]. Although a basic tenet of protein folding is that the native state of a protein is at the minimum free energy[4], we show here that both the I and fully unfolded states of αLP are lower in free energy than the native state. Native αLP is thus metastable: its apparent stability derives from a large barrier to unfolding. Consequently, the evolution of αLP has been distinct from most other proteins: it has not been constrained by the free-energy difference between the native and unfolded states, but instead by the size of its unfolding barrier.

The αLP N state is compact, with a well ordered hydrophobic core, and is stable to chemical and thermal denaturation (see ref. 5 and below). In contrast, the I state has some secondary structure but little or no tertiary structure[1]. Gel filtration[1] and analytical ultracentrifugation (Fig. 1a) indicate that I is a greatly expanded monomer. Not surprisingly, I is temperature-sensitive[1] and it is stabilized by less than 1 kcal mol^{-1} relative to the unfolded state 'U' (Fig. 1b). The conventional view of the forces that drive protein folding[4] would place N at the minimum Gibbs free energy. In principle, however, it is possible for a protein to function even if its native state is not at the free-energy minimum, provided that the kinetic barrier is sufficient to prevent unfolding over the protein's functional lifetime. We now demonstrate that this is the case for αLP.

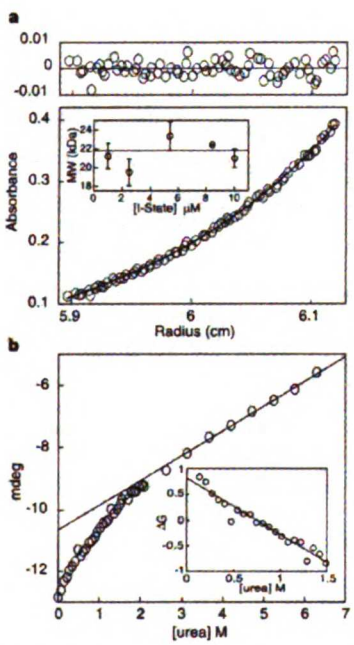


Figure 1 Characterization of the αLP I state. **a**, Sedimentation equilibrium analysis of 10 μM αLP I state with monomer fit and residuals. As shown in the inset, the I state remains monomeric, with an average M_r of 21.8K over at least a 10-fold concentration range. **b**, Urea denaturation of the I state followed by its circular dichroism signal at 225 nm. Inset shows the free energies for the I ↔ U transition as a function of urea, calculated using the displayed linear unfolded baseline and by assigning the I-state baseline to the 0 M urea value. The calculated $\Delta G_{I\leftrightarrow U}$ at 0 M urea must be considered a maximum value.

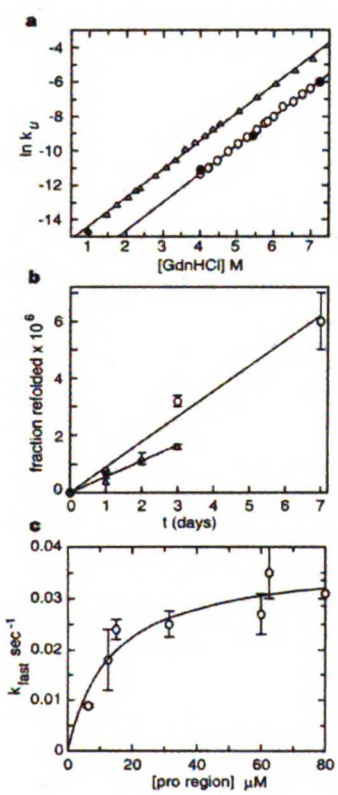


Figure 2 The rates of αLP unfolding and refolding. **a**, The logarithm of the unfolding rates at 4 °C (circles) and 25 °C (open triangles) as a function of denaturant (guanidinium (Gdn) hydrochloride) concentration. Open symbols denote the rates of tertiary structure loss of S195A observed using tryptophan fluorescence; the filled triangle represents the rate of activity loss of the wild-type protease; filled circles show the rates of secondary structure loss in S195A monitored by circular dichroism at 225 nm. **b**, Pro-region-independent folding of the I to N states as a function of time at 4 °C (circles) and 25 °C (triangles). **c**, The concentration dependence of the fast phase rates of pro-region-catalysed I to N folding.

The large folding barrier prevents equilibration of I and N on any practical time scale; thus the relative stability of these states is addressed indirectly through the rates of folding (k_f) and unfolding (k_u). At 25 °C, unfolding of the protease in 1 M guanidinium-HCl can be monitored by loss of activity. With increasing denaturant concentration, measurements of wild-type protease unfolding are complicated by autolysis, so we used a variant of the protease in which the active-site serine is replaced by alanine (Ser 195 → Ala, S195A, where the residue numbering for αLP is based on homology to chymotrypsin[3]) to measure the rates of unfolding as a function of denaturant at 4 °C and 25 °C. During unfolding, secondary and tertiary structure are lost simultaneously and the spectroscopic data are best fitted by a single exponential function of time. The observed rates are plotted as a function of guanidinium-HCl concentration in Fig. 2a. By linear extrapolation to zero concentration of guanidinium, the rate of unfolding is $1.8 \times 10^{-8} \pm 0.2 \times 10^{-8}\,s^{-1}$ at 4 °C ($t_{1/2} = 1.2$ y) and $1.2 \times 10^{-7} \pm 0.6 \times 10^{-7}\,s^{-1}$ at 25 °C ($t_{1/2} = 70$ d). Urea denaturation at 4 °C yields the same extrapolated rate within error (data not shown), supporting the use of a linear extrapolation to determine k_u.

To measure k_f, we incubated solutions of I at 4 °C and 25 °C and assayed aliquots for N-state protease activity as a function of time. The very small fraction of I that refolds during the incubation was detected with an assay of higher sensitivity than that used previously[1] (Fig. 2b). Based on the concentration of refolded N and the total protease concentration, the initial rate of folding k_f was found to be $1.18 \times 10^{-11} \pm 0.06 \times 10^{-11}\,s^{-1}$ at 4 °C ($t_{1/2} > 1{,}800$ y) and $6.0 \times 10^{-12} \pm 0.4 \times 10^{-12}\,s^{-1}$ at 25 °C ($t_{1/2} > 3{,}600$ y). These rates are significantly slower than the rates of unfolding. The equilibrium free energy, calculated from the ratio of k_f to k_u, favours I by 4.0 ± 0.1 kcal mol^{-1} at 4 °C and by 5.9 ± 0.5 kcal mol^{-1} at 25 °C. Therefore, the I state of αLP, not the N state, is at the minimum free energy over a broad range of temperatures.

αLP folds efficiently to its metastable native conformation only with the assistance of its pro-region. The catalysed folding reaction has two time constants, with the population of the fast phase being roughly twice that of the slow phase. The slow phase, which is essentially independent of pro-region concentration, has a time constant of ~200 s (data not shown), and may be a consequence of proline isomerization, as αLP has one *cis* and three *trans* prolines. The fast phase rates are plotted as a function of pro-region concentration in Fig. 2c. From these data, the pro-region is calculated to bind I with 13 ± 5 μM affinity and the folding rate is calculated to be $0.037 \pm 0.004\,s^{-1}$.

With the data shown in Figs 1 and 2, a free-energy diagram for the protease-folding reaction in the presence and absence of pro-region can be constructed (Fig. 3). Remarkably, both I and U are more stable than the folded N state. The pro-region shifts the energetics thermodynamically to favour the N·pro-region complex over the I·pro-region complex, thereby providing the driving force that brings the protease into its native, active conformation. Proteolytic degradation of the pro-region then 'locks' αLP in its metastable active N state. The pro-region is much more than a native-state template. It actively catalyses the folding reaction by stabilizing the folding transition state. The pro-region binds the folding transition state more tightly than I or N, and mutants of pro-region have been identified that preferentially alter this transition-state stabilization[5].

Investigations into the enthalpic and entropic contributions to the free energy difference (ΔG) between I and N provide insight into the unusual stability of the unfolded states of αLP. Because direct measurement of the enthalpy difference (ΔH) between I and N is not possible, we measured ΔH by titration calorimetry using the thermodynamic cycle shown in Fig. 4. At 10 °C, the native conformation is favoured enthalpically by 18 ± 1.5 kcal mol^{-1} over I. Therefore, the stability of I must be entropic in origin.

Comparison of αLP with the homologous but thermodynamically stable protease, chymotrypsin, reveals an intriguing potential source of excess entropy. The αLP sequence contains 16% glycine compared to only 9% in chymotrypsin. Glycines lack a side chain, which increases the number of conformations accessible to unfolded states. The ten additional glycines in αLP are predicted to contribute an additional ~7 kcal mol^{-1} of configurational entropy to the unfolded state compared to that of chymotrypsin at 4 °C (ref. 6). Removal of this entropic source alone would be sufficient to place the N state at the global free-energy minimum for αLP.

An unusually low conformational entropy of the αLP N state may also contribute to increasing the unfolding entropy. Although native states are often dynamic, the N state of αLP is quite rigid: the protease is not readily digested by itself or by other proteases, it has ~40 core amides with $>10^{10}$ protection factors, and the B factors are unusually low (J. Davis, J.L.S. and D.A.A., unpublished results, and ref. 3). Some of the rigidity may arise from the fact that the

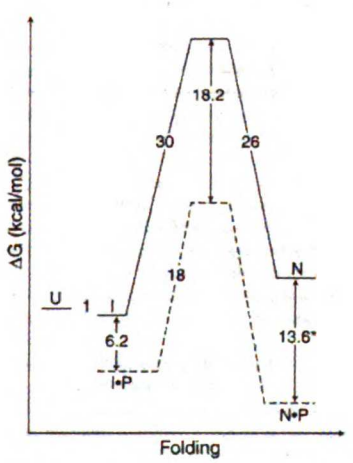


Figure 3 Free energy of αLP folding with and without its pro-region at 4 °C. Using the pro-region I-state binding constant (Fig. 2c), the pro-region N-state inhibition constant[17] and the rates of folding and unfolding (Fig. 2a, b) in combination with transition state theory, the free-energy profile of αLP folding is drawn. Dashed lines denote energetic changes resulting from the addition of pro-region. The free-energy difference between I and U is not precisely known but is estimated to be less than ~1 kcal mol^{-1} (see Fig. 1b). Asterisk indicates that the pro-region binding of the native state was measured at 25 °C and pH 8.

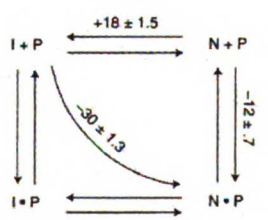


Figure 4 Dissection of the free energy difference between I and N. The enthalpy difference between I and N was determined at 10 °C using the outlined thermodynamic cycle. The enthalpies of I + P to N·P and N + P to N·P were measured. The total enthalpy change around the cycle must equal zero. Therefore, the enthalpy of the N- to I-state transition is $+18 \pm 1.5$ kcal mol^{-1} at 10 °C.

loops in αLP are generally shorter and are likely to be less flexible than those found in chymotrypsin. Many of the glycines unique to αLP are also found in these loops and so may reduce the entropy of the N state while increasing the configurational entropy of the I state.

The rigidity of the native state of αLP may be the product of evolutionary pressures to suppress autolysis and thereby extend the lifetime of the protease, which provides nutrients for its host. The unfavourable and presumably highly cooperative unfolding barrier of αLP would be expected to reduce sources of proteolytic sensitivity such as local breathing motions and/or partial unfolding of the native state. Similarly large kinetic barriers are likely to be present in most extracellular bacterial proteases; virtually all are synthesized with pro-regions and, where examined, the pro-regions catalyse protease folding[7–9]. There is a strong correlation between high glycine content and the presence of a pro-region. As the average glycine content of αLP and other chymotrypsin-like proteases made with pro-regions is 18%, compared with 9% in family members that do not contain a pro-region, incorporating more glycines may be one mechanism by which large kinetic barriers and pro-region catalysts have co-evolved in extracellular bacterial proteases in order to prolong their functional lifetime.

Large barriers have been observed in the folding of several other proteins that are not synthesized with pro-regions. The kinetic barriers of haemaglutinin, luciferase, the serpin PAI-1, and of the prion protein PrP, all function to enhance the stability of one compact native-like state relative to another[10–13]. In contrast, αLP has evolved a kinetic barrier to isolate its native state from unfolded states. This has enabled the protease to optimize its functional properties independently of the thermodynamic consequences, resulting in a kinetically stable but thermodynamically unstable native state. □

Methods

Protein expression and purification. αLP and pro-region were prepared according to published protocols[14]. The construct for the inactive mutant (S195A) has been described[14]. Expression and purification of S195A was as described[15], except that the medium contained 1.5% yeast extract, 1% NaCl, 60 mM ACES, pH 6.3, and induction was at 12 °C. Pro-region was digested from the non-covalent S195A–pro-region complex with trypsin-coupled beads at pH 7 or with pepsin at pH 3. The S195A X-ray structure is virtually identical to that of wild-type protease (S.S.J., S. Rader and D.A.A., unpublished results).

Analytical ultracentrifugation. Sedimentation equilibrium experiments at 4 °C were carried out on I state in 10 mM potassium acetate, pH 5, on a Beckman Model XLA analytical ultracentrifuge at speeds of 18,000 and 22,000 r.p.m., scanning at 230 nm and 280 nm. Data were evaluated using XL-A data analysis software and fitted best at all concentrations to the single non-associating species model. Global analysis of nine data sets gave a relative molecular mass (M_r) of 21.7K with a 95% c.i. of ±0.9K.

Urea denaturation. The stability of I state was measured as a function of urea at 225 nm in an AVIV 62DS circular dichroism spectropolarimeter at 4 °C. 3.5 μM I state in 10 mM potassium acetate at pH 5 was titrated with 4 M and 8 M urea. The unfolded baseline was calculated by a linear fit of the 3–6 M urea data.

Wild-type protease unfolding. 4 μM protease in 0.98 M guanidinium-HCl, 10 mM potassium acetate, pH 5, was incubated at 25 °C. Activity was measured 13 times on duplicate aliquots according to published protocols over 80 days[1].

S195A unfolding. Reactions contained S195A in 10 mM potassium acetate, pH 5, at concentrations of 0.1–1.75 μM (fluorescence) or 7 μM (circular dichroism). Fluorescence measurements were made in a 8100SLM-Aminco fluorimeter. Circular dichroism spectra were monitored on a Jasco J-710 spectropolarimeter. The error in the 25 °C rate constant was determined from the difference between the linear extrapolation and denaturant binding model[16] fits.

Refolding from the I state to the N state. 2.4–6.5 μM I state in 10 mM potassium acetate, pH 5, was incubated at 4 °C and 25 °C. Initially and at each timepoint, 3-ml aliquots of I state were treated with 50 μg pepsin at pH 2.5, which digests the I state but leaves native protease intact, then concentrated. The protease was assayed by following the absorbance change at 324 nm in a solution containing 1 mM succinyl-Ala-Ala-Pro-Ala-thiobenzyl ester (Enzymes Systems Products), 250 μM aldrithiol-4 (Sigma), 2.5% DMSO and 0.1M Tris, pH8. Addition of 1 μM pro-region inhibited the observed activity, indicating that activity is due to αLP N state. The concentration of native protease, $[N]$, was determined from a standard curve. The detection limit is ~6.0 fmolar. Intact I state, $[I]_0$, was measured at each timepoint, as it decreases as a function of time owing to I-state proteolysis by refolded N.

Data from three experiments were fitted simultaneously to determine k_f using data in which the amount of intact I state was within 10% of the starting value. Calculations are based on a reversible equilibrium, $I \leftrightarrow N$. The rate equation for this equilibrium is $[I]_t/[I]_0 = [1/(k_f + k_u)] \times \{k_f\{\exp -(k_f + k_u)t\} + k_u\}$, where the concentration of I state at time t is given by $[I]_t = [I]_0 - [N]_t$ assuming $[I]_t = [I]_0$ at $t = 0$. As both $k_f t$ and $k_u t$ are small, the Taylor series approximation is applied, yielding: $k_f = ([N]/[I]_0)/t$.

Refolding in the presence of the pro-region. Pro-region-catalysed refolding of the I state was measured at 4 °C in 20 mM potassium phosphate at pH 7.2 (ref. 1). The pro-region:I state concentration ratio was maintained at ≥15:1 and reaction rates were fitted by a five-parameter double exponential. The plot of the fast phase rates as a function of pro-region concentration was analysed by assuming a fast I + P ↔ I·P pre-equilibrium followed by a slow, concentration-independent I·P → N·P transition.

Titration calorimetry. Enthalpies were measured in 20 mM potassium acetate, 42 mM guanidinium-HCl, pH 5, buffer using a Microcal Omega titration calorimeter. Pro-region (~100 μM) was titrated with I or N from 8–12 °C. The enthalpies at 10 °C are the average of nine measurements. Temperatures below 8 °C and above 12 °C were not used owing to low signal-to-noise and complications from pro-region unfolding, respectively.

Received 15 May; accepted 30 July 1998.

1. Baker, D., Sohl, J. L. & Agard, D. A. A protein-folding reaction under kinetic control. *Nature* **356**, 263–265 (1992).
2. Silen, J. L. & Agard, D. A. The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. *Nature* **341**, 462–464 (1989).
3. Fujinaga, M., Delbaere, L. T., Brayer, G. D. & James, M. N. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **184**, 479–502 (1985).
4. Makhatadze, G. L. & Privalov, P. L. Energetics of protein structure. *Adv. Protein Chem.* **47**, 307–425 (1995).
5. Peters, R. J. *et al.* Pro region C-terminus: Protease active site interactions are critical in catalyzing the folding of α-lytic protease. *Biochemistry* **37**, 12058–12067 (1998).
6. D'Aquino, J. A. *et al.* The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25**, 143–156 (1996).
7. Baker, D., Shiau, A. K. & Agard, D. A. The role of pro regions in protein folding. *Curr. Opin. in Cell Biol.* **5**, 966–970 (1993).
8. Eder, J. & Fersht, A. R. Pro-sequence-assisted protein folding. *Mol. Microbiol.* **16**, 609–614 (1995).
9. Bryan, P. *et al.* Catalysis of a protein folding reaction: mechanistic implications of the 2.0 Å structure of the subtilisin–prodomain complex. *Biochemistry* **34**, 10310–10318 (1995).
10. Clark, A. C. *et al.* Kinetic mechanism of luciferase subunit folding and assembly. *Biochemistry* **36**, 1891–1899 (1997).
11. Chen, J. *et al.* A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc. Natl Acad. Sci. USA* **92**, 12205–12209 (1995).
12. Harrison, P. M., Bamborough, P., Daggett, V., Prusiner, S. B. & Cohen, F. E. The prion folding problem. *Curr. Opin. Struct. Biol.* **7**, 53–59 (1997).
13. Wang, Z., Mottonen, J. & Goldsmith, E. J. Kinetically controlled folding of the serpin plasminogen activator inhibitor 1. *Biochemistry* **35**, 16443–16448 (1996).
14. Sohl, J. L., Shiau, A. K., Rader, S. D., Wilk, B. J. & Agard, D. A. Inhibition of α-lytic protease by pro region C-terminal steric occlusion of the active site. *Biochemistry* **36**, 3894–3902 (1997).
15. Mace, J. E. & Agard, D. A. Kinetic and structural characterization of mutations of glycine 216 in α-lytic protease: a new target for engineering substrate specificity. *J. Mol. Biol.* **254**, 720–736 (1995).
16. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).
17. Baker, D., Silen, J. L. & Agard, D. A. Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins* **12**, 339–344 (1992).

Acknowledgements. We thank P. Walter and P. Peluso for use of their fluorimeter, S. Marquese for use of her Aviv CD and titrator, Y.-R. Yang for assistance with analysis of data from analytical ultracentrifugation, N. Sauter for assistance with figures, and A. Derman, M. Gustafsson, T. Mau, A. Shiau and H. S. Chan for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute. S.S.J. is supported by a Howard Hughes Medical Institute predoctoral fellowship. J.L.S. was supported by NIH training grants.

Correspondence and requests for materials should be addressed to D.A.A. (e-mail: agard@msg.ucsf.edu).

Chapter 2. Characterization of the αLP Intermediate.

Introduction

From the previous chapter, the intermediate state of αLP emerged as a key feature of the protein folding landscape. This molten globule-like conformation, despite displaying high sensitivity to proteolysis, temperature and denaturant, is actually at the global free energy minimum. The excess stabilization over the native state is likely due to an unusually large entropy difference between it and the native state, resulting from the high number of glycines in the sequence increasing the entropy of the intermediate state and possibly decreasing the entropy of the native state. However, many questions remained, which prompted us to further explore the nature of this conformation.

How does it compare to other protein folding intermediates in terms of its structure and energetics? Is it possible to more accurately determine its stability? How unfolded is it relative to the native state? Our ability to address these questions was hindered by the practical limitations of working in a very narrow temperature range at concentrations below 10 μM, due to the extreme sensitivity of the intermediate state. Therefore this chapter outlines the preliminary progress we made in answering these questions, in addition to efforts made to stabilize the intermediate in order to facilitate future experiments.

Methods

ANS binding

ANS was made up at 10 mM in EtOH. then diluted to 1.5 mM in 10 mM KOAc pH 5.0, and used at 50 μM in the unfolding reactions. For wavelength spectra, ANS

fluorescence was measured after excitation at 380 nm by following emission between 450 – 650 nm. ANS binding at each urea concentration was followed by monitoring fluorescence emission at 475 nm and 482 nm for 10 seconds after excitation at 380 nm.

Equilibrium Urea denaturation

Stock solutions of 2 or 9 μM Intermediate were made in 0, 3 and 8 M urea, 10 mM KOAc pH 5, 50 μM ANS. Following determination of the fluorescence of the initial 0 M urea intermediate sample, aliquots of Intermediate were successively titrated to increment the urea concentration up to 7.9 M urea, and equilibrated for 5 minutes before measurement of the fluorescence. After each titration, equilibration and measurement, an aliquot was removed to accurately determine the urea concentration by refractive index.

Analysis of Urea denaturation

sigmoid + linear function: m1+m2*m0+m3/(1+exp(m4*m0))

shifted linear function: m1+m2*m0+m3/(1+exp(m4*(m0-m5)))

Results

ANS binding as a probe of the Intermediate

Previous work showed that the intermediate has several characteristics of a typical molten globule intermediate, it has 2 ° but no 3 ° structure, and has a monomeric conformation that is greatly expanded over the native state (Baker et al., 1992b). However, the defining characteristic of a molten globule intermediate is its exposure of hydrophobic surface area, which can be determined by the ability to bind the hydrophobic dye ANS (ref). Because ANS binds hydrophobic cavities, its fluorescence is significantly enhanced in the presence of molten globules but not native or completely unfolded states.

A rotation student, Jill Nephew and I examined the ability of αLP in its various conformational states to bind ANS. We demonstrated that there is indeed a greatly enhanced ANS fluorescence signal in the presence of the αLP intermediate sample compared to the unfolded protease in 8 M urea, and the native protease (Figure 2.1).

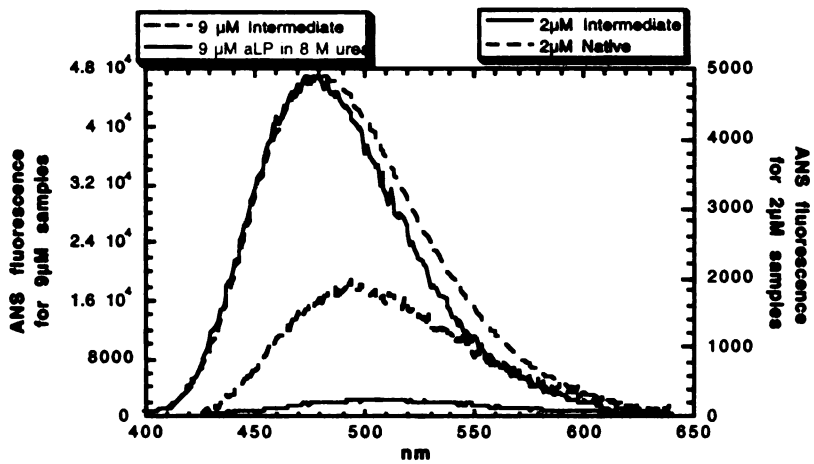


Figure 2.1 ANS binds to the intermediate. ANS Fluorescence is significantly enhanced in the presence of αLP intermediate (red solid 9 μM, dashed 2 μM) compared to unfolded αLP in 8 M urea (9 μM blue) and native αLP (2 μM black).

In addition to confirming the molten globule status of the αLP intermediate, this result provided an alternative avenue for probing the intermediate state besides circular dichroism. Although the intermediate displays a strong circular dichroism signal which was used to monitor the urea denaturation of the intermediate at 4 °C (Sohl et al., 1998), that signal can become problematic at higher temperatures. As the temperature is increased, the ellipticity observed for the intermediate can actually increase (Sohl, 1997) (see Chapter 5). Since a full characterization of the thermodynamics of the intermediate would include probing the effect of temperature, ionic strength, pH, etc., we wanted to

avoid relying only on circular dichroism. Using ANS fluorescence also offered the possible advantage of higher sensitivity, so that we could use less protein to examine intermediate stability under a large number of conditions.

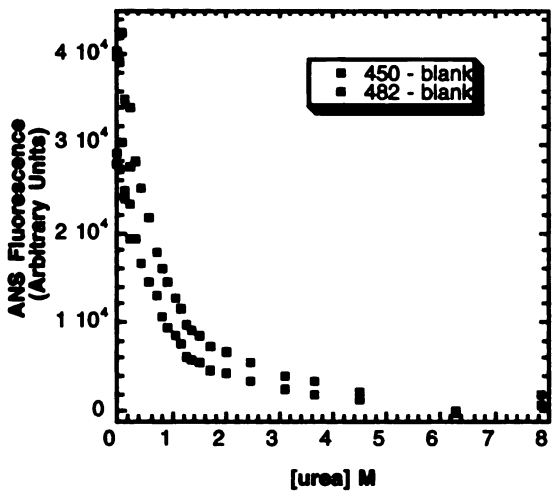


Figure 2.2 Urea denaturation of intermediate followed by ANS fluorescence. ANS fluorescence at 450 nm or 482 nm decreases as the urea concentration increases.

Figure 2.2 shows the urea denaturation of the intermediate followed by change in ANS fluorescence at the wavelength of maximum emission and a wavelength on the shoulder. As with the denaturation of the intermediate by CD, the data do not reach a plateau at a "folded" baseline. With the CD data, the value of the ellipticity for the folded intermediate could be estimated based on the ellipticity for the same concentration of the native state.

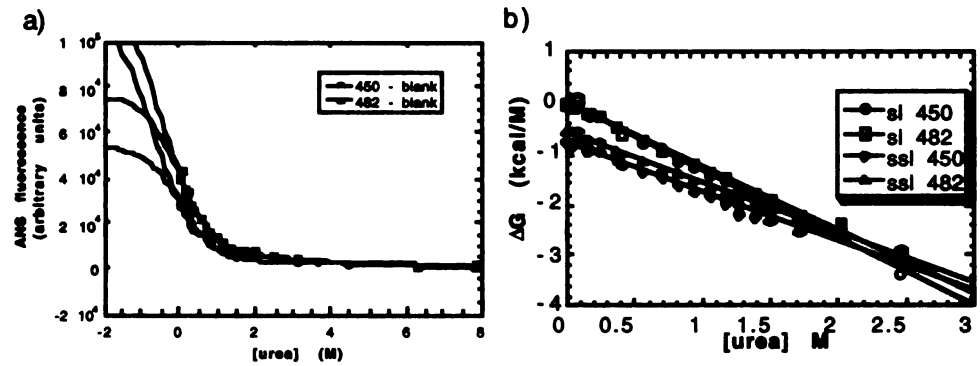


Figure 2.3 Fits of urea denaturation of intermediate. a)ANS fluorescence vs. urea concentration. b) ΔG_{I-U} vs. urea.

The ANS data can be fit by using variations of a sigmoid + linear function to fit the data (Figure 2.3a) (see Methods). The free energy of unfolding the intermediate at each urea concentration can be calculated based on taking either of those fit baselines as the native baseline. As the plot of ΔG_{I-U} vs. urea in Figure 2.3b shows, the estimate of the stability of the intermediate in the absence of urea varies from near 0 kcal/mol if determined from the sigmoid + linear fit, to -0.7 kcal/mol based on the shifted sigmoid + linear fit. The slope of the urea dependence of ΔG_{I-U} varies from -1.3 to -1 kcal/mol•M.

Those slopes correspond well with the slope from Julie Sohl's urea denaturation of the intermediate monitored by CD (Sohl et al., 1998),which was about -1 kcal/mol•M. However the CD measurements gave a maximum estimate for the stability of the intermediate (ΔG_{I-U}) of 1 kcal/mol. As mentioned above, that estimate was based on assuming that the folded baseline of the intermediate ellipticity would not exceed that of the folded native protease and therefore assigning the native protease value as the folded intermediate value. Based on the fact that intermediate ellipticity can in some cases

exceed the native ellipticity, that assumption may not be valid. That could account for the difference between the stabilities estimated from CD vs. ANS binding.

Stabilizing the Intermediate

Having established that ANS binding can be used to monitor the stability of the intermediate, our next goal was to test various small molecule stabilizers for their ability to increase the stability of the intermediate. This would be helpful for determining a more accurate stability from a complete equilibrium denaturation curve including a native baseline. In addition, a more stable intermediate might prove more amenable to concentration for structure determination techniques such as small angle x-ray diffraction and hydrogen exchange studies monitored by NMR.

Since the late 1970s there have been numerous reports of stabilization of proteins by small molecules such as sucrose, ethylene glycol, and glycerol (Lee and Timasheff, 1981), (Gekko and Timasheff, 1981), (Arakawa and Timasheff, 1982), (Timasheff, 1995). Recently TMAO (Trimethylamine-N-oxide) has been shown to cause a huge increase in the stability of two proteins which are normally unfolded under physiological conditions (reduced and carboxyamidated RNase T1, and a destabilized mutant of staphlococcal nuclease) (Baskakov and Bolen, 1998). The mechanism of stabilization has been attributed to both to effects on solvation as well as direct interactions with the native protein (Baskakov and Bolen, 1998). Depending on the mechanism of stabilization, various stabilizers may have the effect of stabilizing molten globule states of proteins over the unfolded states, or may stabilize the native state over both molten globule and unfolded states.

As a first pass, we decided to examine ethylene glycol, glycerol, sucrose and TMAO for possible stabilizing effects. However, when we tested the effect on ANS fluorescence of just those molecules alone in 0.5 M urea, we found that they all enhanced ANS fluorescence significantly (Figure 2.4). In all cases the effect was greater than that of 2 μM Intermediate alone (see Fig. 2.1). In fact, the ANS fluorescence in glycerol and ethylene glycol was nearly half of the ANS fluorescence in the presence of 9 μM Intermediate, which is near the maximal workable concentration of Intermediate.

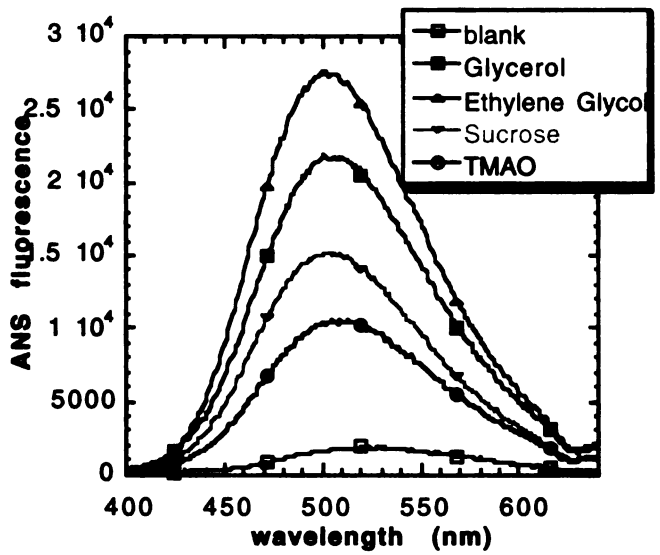


Figure 2.4. Effect of stabilizers alone on ANS fluorescence. 10 mM ANS in 0.5 M urea alone (open squares). with 38% glycerol (closed squares), 38% ethylene glycol (upright triangles), 38% sucrose (downward triangles), or 2.6 M TMAO (circles).

Because TMAO had the smallest effect on ANS fluorescence alone, we decided to pursue it as a possible stabilizer. A comparison of the normalized ANS fluorescence in the presence of 4.5 μM Intermediate alone or with 2.6 M TMAO indicated that there was greater than a threefold increase in the ANS fluorescence enhancement with TMAO

present (Figure 2.5). This suggested that TMAO was able to stabilize the "folded" molten globule form of the intermediate.

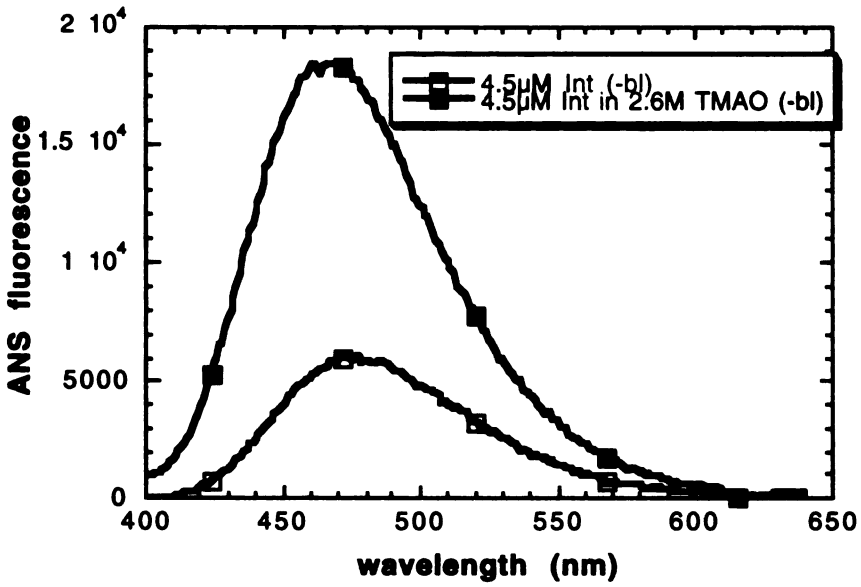


Figure 2.5 Effect of 2.6 M TMAO on intermediate.

We then followed ANS fluorescence during titration of the intermediate with TMAO in order to determine TMAO concentration that leads to the maximal stability of the intermediate (Figure 2.6). This would allow us to then perform the urea titration at that concentration of TMAO in order to observe a full equilibrium denaturation curve, including a native baseline for the intermediate. By following the urea denaturation of the intermediate at a series of TMAO concentrations where the complete transition is observable, we could extrapolate ΔG_{I-U} to zero TMAO in order to determine a more accurate ΔG_{I-U}.

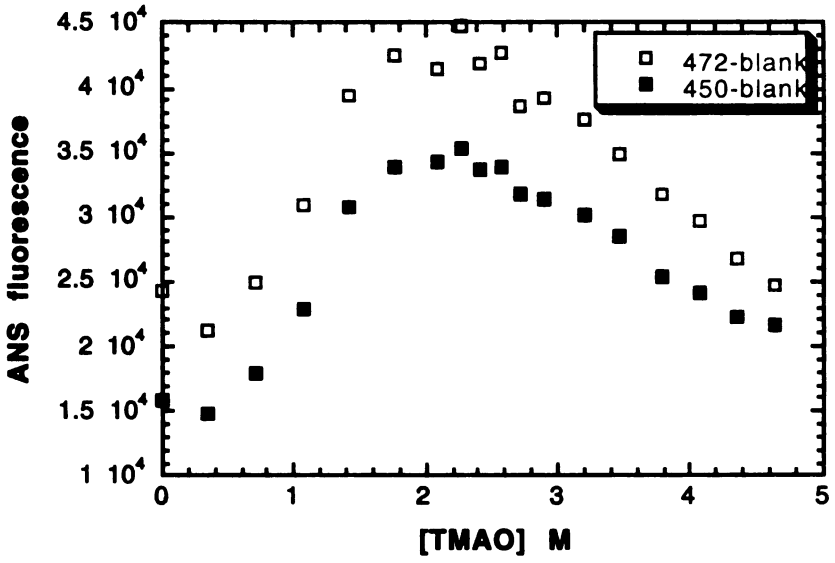


Figure 2.6 TMAO titration of intermediate.

However, although we observed increasing enhancement of ANS fluorescence with 0 – 2.2 M TMAO, where the fluorescence plateaued, and then started to decrease. Because the intermediate sample with the final concentration of 4.8 M TMAO aggregated after two days, we reasoned that high TMAO concentrations were causing association of the intermediate. In fact, when we tried to make stock solutions or 9 μM Intermediate even in 2.2 M TMAO in order to titrate with urea, the protein immediately aggregated.

Since we didn't see aggregation in the samples of intermediate and 2.6 M TMAO in 0.5 M urea, it's likely that the urea was able to suppress aggregation. However, no systematic observation of the change in those samples over time were made, and it is possible that aggregation also eventually occurred in them. Since Jill's rotation came to an end without a clear understanding of whether TMAO is actually stabilizing or simply

increases the hydrophobicity of the intermediate in a way that then leads to aggregation, no further progress with stabilizing the intermediate has been made.

Discussion

The experiments described above have led to a better general understanding of the structure of the αLP intermediate and its position in the folding reaction. Demonstration of ANS binding confirms that the αLP intermediate does fit into the general category of "molten globule" protein folding intermediate, as expected based on its other characteristics (expanded radius, 2° but no 3° structure).

The best estimate of stability based on using ANS binding to follow equilibrium urea denaturation is between -0.6 and 0 kcal/mol. This suggests that the analysis of the unfolding data by CD is overestimating the stability by 0.4 – 1 kcal/mol. However, since there is no true estimate of the folded baseline for the denaturation followed by ANS binding,, the CD estimate is still the best estimate.

The m value for the urea denaturation of the intermediate followed with ANS binding closely matches the value of 1100 cal/mol•M from the CD measurements. m values have been shown to be proportional to the amount of surface area exposed on unfolding (Myers et al., 1995). Based on the calculated m value for complete unfolding from N to U, which is 2100 cal/mol•M (see Chapter 4), the I to U m value, at ~1100 cal/mol•M indicates that the intermediate has buried about 50% of the surface exposed in the unfolded state. This sets up the following reaction coordinate for the folding of αLP based on burial of surface:

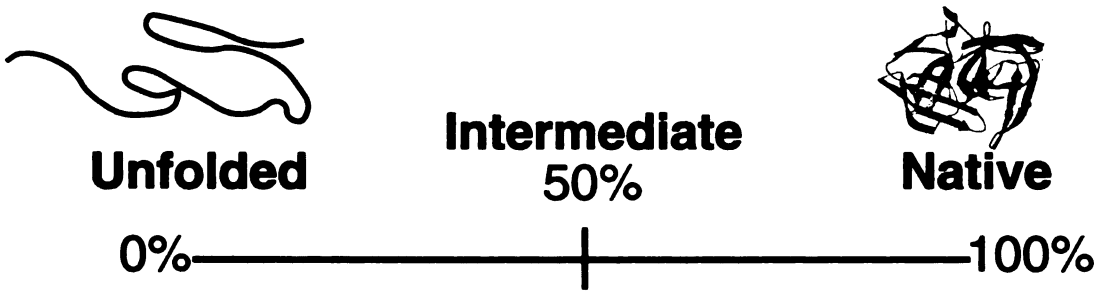


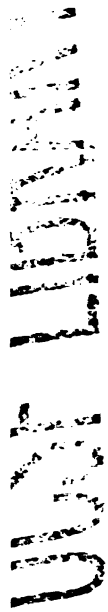
Figure 2.7. αLP folding reaction coordinate based on surface area burial.

Future Directions

There are many remaining questions about the intermediate that deserve to be addressed. Determining the accurate stability is still a reasonable goal. It appears that because of the high background of ANS fluorescence in the presence of stabilizers alone, using CD to measure the effect of stabilizers on intermediate stability is a more viable option, despite the higher concentration of intermediate required. A lower concentration of stabilizer (less than the 38% used in these attempts) could also be tried. It is likely that one of the other weaker stabilizers may stabilize the intermediate without leading to aggregation, as is the case with TMAO, which is known to be quite a strong stabilizer (Baskakov and Bolen, 1998). A more detailed description of the thermodynamic parameters involved in stabilizing the intermediate could be determined by following the temperature dependence of the intermediate unfolding by urea in a limited temperature range (4 – 25 °C), since complications arise as the temperature is increased (Chapter 5).

Due to the fact that the intermediate is 50% folded with respect to burial of surface area, it is very likely that protection from hydrogen exchange could be detectable. If it becomes possible to reach NMR level concentrations through the use of stabilizers,

then it would be very exciting to know the location of protected residues in terms of the native structure. However, mass spectrometry could be attempted at low concentrations to monitor the extent of hydrogen exchange, and a combination of mass spectrometry and proteolysis could be tried to determine the identity of protected regions (Miranker et al., 1996).



Chapter 3. Linear Extrapolation, Denaturant Binding, and Solvent Transfer Models are inadequate for analysis of αLP unfolding kinetics over a wide denaturant and temperature range.

Introduction

Knowledge of the stability of proteins in their native and intermediate conformations, as well as the heights of the kinetic barriers between them is central to understanding protein folding and design. The use of denaturants to induce unfolding is one of the most common methods applied in the investigation of protein stability and barriers (Pace, 1986), (Chen et al., 1989). The exact mechanism by which denaturants promote unfolding is still a matter of considerable debate, but is likely to involve improved solvation of the hydrophobic surface exposed on unfolding, through effects on the solvent structure and/or effects from direct interaction of the denaturant molecules with the protein.

Typically the native state is assumed to be in equilibrium with the unfolded state (Pace, 1986), (Myers et al., 1995).

$$N \Leftrightarrow U \qquad (1)$$

$$K = [U]/[N] \qquad (2)$$

The free energy of unfolding the native state is calculated from the equilibrium constant:

$$\Delta G_U = -RT\ln[U/]/[N] \qquad (3)$$

(A positive ΔG_U denotes an unfavorable free energy.)

For most proteins under physiological conditions, the equilibrium is so far in favor of the native state that it is impossible to detect the unfolded state in order to determine the ratio of their concentrations. Therefore moderate to high concentrations of denaturant are used to destabilize the native state enough to detect the unfolded state. The free energy of

unfolding in the presence of denaturant (ΔG_U) is measured in the transition region where both N and U are populated. Typically K can only be measured between 0.1 to 10, corresponding to a change in ΔG_U of ± 2.8 kcal/mol (Creighton, 1993). ΔG_{H2O}, the free energy of unfolding in the absence of denaturant, is then determined from extrapolation of the measured ΔG_U in the experimentally accessible denaturant range to zero denaturant using methods described below.

Similarly, in the analysis of barriers to protein unfolding, the native and transition states are considered to be in equilibrium (Chen et al., 1989).

$$N \underset{\nu}{\overset{k_u}{\Leftrightarrow}} Ts \qquad (4)$$

$$\nu = k_B T/h \qquad (5)$$

(k_B = Boltzmann's constant, h = Planck's constant)

The rate of unfolding k_u can be measured by following the kinetics of structure loss at a series of denaturant concentrations. The dependence of ln k_u on denaturant can then be analyzed to determine ln k_u in the absence of denaturant, or a $\Delta G^{\ddagger}_{U,den}$ can be calculated for each denaturant concentration based on transition state theory (Chen and Matthews, 1994):

$$K = k_u/\nu \qquad (6)$$

$$\Delta G^{\ddagger}_U = -RT\ln(k_u/\nu) \qquad (7)$$

and extrapolated to determine $\Delta G^{\ddagger}_U$.

Since it is usually only possible to measure unfolding in a range of denaturant concentrations far from zero denaturant, various methods of analysis have been applied to extrapolate to the ΔG_U in the absence of denaturant based on the dependence of the

measured ΔG_U on denaturant concentration. These methods are also applied to the extrapolation of $\Delta G^{\ddagger}_U$. The simplest method is the linear extrapolation model:

$$\Delta G_U = \Delta G_{H2O} - m[D] \qquad (8)$$

with ΔG_{H2O} as the intercept and m as the slope, which was developed from the empirical observation that ΔG_U appeared to have a linear dependence on denaturant (Aune and Tanford, 1969), (Pace, 1986). (D is most often treated as straight M concentration or transformed to denaturant activity in terms of molar ion activity(Aune and Tanford, 1969), (Pace, 1986), or hydrophobic solvation energy (Parker et al., 1995). The m value has traditionally been interpreted as an indicator of the amount of hydrophobic surface area exposed in the unfolding transition, and was found to be proportional to the total change in exposed surface area and ΔC_p of unfolding (Myers et al., 1995) for a set of 45 proteins with known structures.

This method has enjoyed the greatest popularity, and in some experimental cases has appeared justified. Linear extrapolations from both urea and GdnHCl yield similar values of ΔG_{H2O} for lysozyme (Greene and Pace, 1974) and chymotrypsin (Greene and Pace, 1974), (Bolen and Santoro, 1988). A combination of GdnHCl and thermal unfolding data for thioredoxin by DSC was found to be linear from 0 – 3 M and match the linear urea extrapolation when the chloride concentration was kept fixed (Santoro and Bolen, 1988). A thermodynamic cycle using pH titration of native and unfolded states demonstrated the validity of LEM for urea data of RNase A (Yao and Bolen, 1995) and GdnHCl data for HPr (Nicholson and Scholtz, 1996). Hydrogen exchange (HX) measurements in the absence of denaturant confirmed the linearly extrapolated values for RNAse H (Chamberlain et al., 1996), NTL9 (Kuhlman et al., 1998), and lambda

repressor (Huang and Oas, 1995), (Huang and Oas, 1996). However, in many cases, such tests are not applied and when they are, these are usually used only under one set of conditions, and then assumed to apply under other conditions.

In addition, there are also many reports of different estimates of ΔG_{H2O} obtained from unfolding in urea vs. GdnHCl. For metmyoglobin (Gupta et al., 1996) and FABP (Ropson et al., 1990), linear extrapolations yield higher values of ΔG_{H2O} from urea than GdnHCl, while for RNAse A (Greene and Pace, 1974), (Yao and Bolen, 1995)the linearly extrapolated GdnHCl value is higher than from urea. With the application of HX to measure unfolding at conditions of very low and absent denaturant, and the improvement in the range and quality of data obtained through traditional spectroscopic methods, it has become increasingly possible to detect experimental evidence for curvature in the dependence of ΔG_U and ln k_u on denaturant for some proteins. A combination of hydrogen exchange data and GdnHCl and temperature CD melts of Protein L shows a distinct curvature leading to the HX measured ΔG_{H2O} higher than the linearly extrapolated value from the higher concentration CD data alone (Yi et al., 1997). The dependence of ΔG_U on urea from urea and thermal denaturation for barnase displays upward curvature as well (Matouschek et al., 1994). Additionally, theoretical calculations based on the observation of curvature in the dependence of the free energies of transfer of hydrophobic amino acids from water to different concentrations of GdnHCl predict a curvature in the GdnHCl dependence (Alonso and Dill, 1991), (Makhatadze, 1999).

Despite the wealth of evidence against the universality of the LEM approach, values of the free energy and rates of unfolding determined using the LEM are often used

to compare mutants and evaluate stabilities of designed proteins. In addition, the variation of both ΔG_U and rates of unfolding with temperature and pH conditions are used to analyze thermodynamic parameters of protein unfolding reactions. Therefore it is of general interest to the field to thoroughly evaluate the methods for extrapolating unfolding measurements made in denaturant, and to understand the effect of denaturants on both the solvent and the protein in order to properly remove all effects which are not related to the protein.

As mentioned above, the chief limitation in obtaining denaturant dependent unfolding data for most proteins is that the transition region in which it is possible to measure a mixture of native and unfolded populations is very narrow and far from zero denaturant. Measurement of the rates of protein unfolding in denaturant are also limited to denaturant concentration ranges where $K > 1$. V shaped chevron plots of ln k_u vs. denaturant are observed where below the $C_{1/2}$ for unfolding, the refolding rate becomes faster than the unfolding rate, and therefore it's not possible to follow protein unfolding kinetics to very low denaturant concentrations (Jackson and Fersht, 1991). As curvature is most likely to occur at lower denaturant concentrations, there is often not enough data to suggest that additional parameters warranted beyond linear, so that is the standard analysis applied in both equilibrium and kinetic unfolding.

The kinetically stable protein αLP, offers a unique advantage in measuring the unfolding rate in denaturant. The rate of unfolding (k_u) is slower than the rate of refolding (k_f) by three orders of magnitude (Sohl et al., 1998). This means that the contribution of k_f to k_{obs}, which is a relaxation rate from both the rate of k_f and k_u, is small enough to be neglected even at low to zero denaturant concentrations. Therefore the

measured k_{obs} over the entire denaturant range can be treated as solely from the rate of unfolding.

Experiments reported here taking advantage of the broad range of temperatures 4 – 60 °C and denaturant concentrations (GdnHCl: 0 – 7.25 M, urea 0 – 8 M) over which k_u is accessible reveal not only a distinct curvature in the dependence of the unfolding rate on GdnHCl and urea, but also a heat capacity to that curvature. These high quality data afford a highly accurate description of the curvature and its change with temperature, leading to a more complete analysis of the thermodynamics of protein-denaturant interactions than before. This has allowed us to test the different models for extrapolation in order to develop a method that describes the solvation and temperature effects of denaturant on unfolding in a physically interpretable manner.

Experimental Methods

Materials

GdnHCl and Urea (ultrapure) were from ICN (Cleveland, OH, U.S.A.) All other chemicals were from Fisher.

SA195 αLP expression

D1210 *E. coli* transformed with the bicistronic pro region SA195 construct (Sohl et al., 1998) were grown at 37 °C in 4.5 – 6 L either LB media or LXY media (15 g yeast extract, 10 g NaCl/ liter) buffered at pH 6.8 with 60 mM ACES to an OD_{600} of 1.8 (LB) or 2.5 (LXY), shifted to 12 °C for two hours, then induced with 0.1 mM IPTG.

SA195 αLP purification

Supernatant from the cells harvested 6 – 7 days post induction was diluted 4 – 5-fold with dH_2O and brought to pH 4.5 – 5. 50 mls S-Seph beads were added and mixed overnight. Beads were collected and washed with 10 volumes 10 mM NaOAc pH 5.0, then 5 volumes either 10 mM Glycine, pH 9.6 or 10 mM Hepes, pH 8.4. SA195 was eluted in a noncovalent complex with the pro region using 10 mM Glycine pH 9.6, 200 mM NaCl. Incubation with trypsin coupled beads at pH 7.5 or 0.1 mg/ml pepsin at pH 3 for 1 – 12 hours followed by dialysis against or 1:4 dilution with 10 mM NaOAc pH 5 digested the pro region. Post protease treated material was loaded onto either Pharmacia MonoSHR10/10 HPLC and eluted with 10 mM NaH_2PO_4 10 – 500 mM NaOAC, pH 7.2, or Vydac VHP81010 and eluted with 10 – 500 mM NaOAc pH 5. Final protein was judged pure as determined by silver stained gel, mass spectrometry and N-terminal sequencing. The x-ray structure of SA195 is virtually identical to that of the wild type (S.S.J., S.D. Rader and D.A.A. (Rader, 1997)).

Kinetic Unfolding Experiments

Unfolding reactions were done in 10 mM KOAc pH 5.0 and initiated by manual mixing to a final SA195 concentration of 0.1 to 1.75 μM for fluorescence or 7 μM for circular dichroism studies. Guanidine denaturation was carried out at 4, 15, 25, 37 50, 60, 70 °C. Refractive index measurements were used to accurately measure the concentration of GdnHCl in each sample (Pace, 1986). Fluorescence experiments were done using an 8100SLM-Aminco Fluorimeter connected to an external thermostat bath. Excitation was at 283 nm and emission at 322 nm. Circular Dichroism experiments were monitored at 225 nm using a Jasco J-715 Spectropolarimeter with a Peltier temperature control.

At 25 °C and 37 °C, unfolding reactions at 0.5 – 1 M GdnHCl were incubated in sealed fluorescence cuvettes in incubators maintained at the appropriate temperatures ± 1 °C and monitored over a period of 4 months.

Methods of Analysis

Analysis of Unfolding Kinetics

Nonlinear regression analysis with the program Kaleidagraph (Abelbeck Software, Reading, PA) was used to obtain the monoexponetial rate constants for unfolding. Applying transition state theory by the method of Chen,et al. (Chen and Matthews, 1994) the activation free energies at each denaturant concentration were determined from the rate constants using *(7)*.

Denaturant Binding Model

The denaturant binding model is based on a physical model of specific binding of denaturant to the protein molecule, with the unfolded state exposing more binding sites, leading to unfolding in the presence of denaturant (Tanford, 1970), (Pace and Vanderburg, 1979), (Makhatadze, 1999). Traditionally this analysis takes the form of

$$\Delta G_U = \Delta G_{H2O} - \Delta nRT \ln (1 + kD), \qquad (9)$$

by assuming that the sites are independent and non interacting and have the same affinity for denaturant whether on the native or unfolded molecule. Δn is the difference in sites between N and U, k is the binding constant to each site, and D can be in terms of *a* (the molar ion activity of the denaturant) or more commonly M (molarity) (Pace, 1986). This model accommodates a changing slope of dependence of ΔG on denaturant (m value in

LEM analysis). It is possible to calculate an "effective m value" based on the fit parameters:

$$\text{"m"} = \Delta nRT \ln (1 + k) \qquad (10)$$

(under standard conditions of 1 M)

This value should report on exposure of surface area in an equivalent manner to the LEM m value.

The nature of the denaturant binding site is unclear. It is presumed that denaturant binds to backbone; it has been suggested that the number of sites is equal to 1/2 the number of peptide bonds in the protein (Pace, 1986). A correspondence was seen between k's determined from fitting denaturant induced unfolding data and calorimetric measurement of the heats of titration of denaturant binding to three proteins (Makhatadze and Privalov, 1992).

Solvent Transfer model

Another model developed by Tanford (Tanford, 1964) accommodates curvature but is based on a different approach for the denaturant effect. The solvent transfer model ascribes the effect of denaturant on the stability of the unfolded state to the increase in solubility of the groups newly exposed to solvent on unfolding. The denaturant dependence is described by the sum of the free energies of transfer for the groups newly exposed to solvent upon unfolding. This was developed based on the thermodynamic cycle:

$$\begin{array}{ccc} & \Delta G_{N\text{-}U} \text{ in } H_20\ (\Delta G_{H2O}) & \\ N \text{ in } H_20 & \Leftrightarrow & U \text{ in } H_20 \\ \Delta G_{tr}\, N \mid & & \mid \Delta G_{tr}\, U \\ N \text{ in denaturant} & \Leftrightarrow & U \text{ in denaturant} \\ & \Delta G_{N\text{-}U} \text{ in denaturant } (\Delta G_U) & \end{array} \qquad (12)$$

From this cycle,

$$\Delta G_U - \Delta G_{H2O} = \Delta G_{tr}\, U - \Delta G_{tr}\, N \text{ (Pace, 1975)} \qquad (13)$$

It is clear that the ΔG_{tr} of groups exposed in both N and U will cancel out, so ΔG_{tr} U - ΔG_{tr} N can be determined from the sum of the total number of groups of type i in the protein times the fractional exposure of each group times the ΔG_{tr} for each group:

$$\Delta G_{tr}\, U - \Delta G_{tr}\, N = \sum \alpha_i n_i \delta g_{tri} \qquad (14)$$

ΔG_{tr} for the peptide group and amino acid sidechains have been determined from studies on the solubility of model compounds in various concentrations of urea and GdnHCl (Nozaki and Tanford, 1970), (Staniforth et al., 1993). Different investigators have chosen to include the peptide group and different sidechains as contributing (ranging from just the hydrophobic sidechains (Nozaki and Tanford, 1970) to all of them (Staniforth et al., 1993). Initially α_i was taken outside of the sum and chosen to be an average value α, to give the observed dependence of ΔG on denaturant concentration, yielding:

$$\Delta G_U = \Delta G_{H2O} + \alpha \sum n_i \delta g_{tri} \text{ (Pace, 1975).} \qquad (15)$$

Recently, Staniforth *et al.* (Staniforth et al., 1993) have explicitly calculated the fractional exposure α_i for the groups exposed during unfolding of the protein under consideration based on the change in accessible surface area (ΔASA) for each buried amino acid sidechain, calculated from the known native crystal structure and a model of the unfolded chain.

$$\alpha_i = (ASA_i \text{ in extended chain} - ASA_i \text{ in native}) / ASA_i \text{ in extended chain} \qquad (16)$$

[illegible]

The number of sidechains of each type in the core is determined from the fractional exposure for each sidechain (α_i) multiplied by the total number of sidechains of that type (N_i):

$$n_i = Ni\ \alpha_i \qquad (17)$$

The sum of those sidechains defines the composition of the solvent excluded core:

$$n = \sum n_i \qquad (18)$$

Using this determination of the core composition of the protein in combination with the model compound data, the solvation energies can then be appropriately summed at each GdnHCl concentration, to give the change in solvation energy for an average internal side chain of that protein.

$$\Delta G_{solv} = (\sum \Delta G_{s,i} n_i)/n \qquad (19)$$

They analyze the denaturant dependence of this average solvation energy using a hyperbolic fit:

$$\Delta G_s(D) = \Delta G_{sm}[D]/(K_{den} + [D]) \qquad (20)$$

This allows them to determine a maximum solvation energy at infinite denaturant concentration ($\Delta G_{s,m}$), and K_{den}, the concentration at which the half maximal solvation energy is reached, for an individual protein. These parameters are then used to analyze the ΔG_U of unfolding with denaturant to determine the ΔG_{H2O} and n, which is the fractional number of buried sidechains exposed on unfolding.

$$\Delta G_U = \Delta G_{H2O} + n\Delta G_{s,m}[D]/(K_{den} + [D]) \qquad (21)$$

We calculated the ΔG_{solv} for an average internal side chain of αLP based on its composition using equations *(16) – (19)*. Table 3.1 shows the composition of the solvent-excluded core based on this analysis. With these n_i's the total n and the $\Delta G_{transfer}$ for

sidechains from model compound data, the free energy of solvation for an average buried sidechain in αLP (ΔG_{solv})was calculated at 1 – 6 M GdnHCl using equation *(20)*. The dependence of αLP 's ΔG_{solv} on denaturant was fit using equation *(21)* (Figure 3.1), yielding parameters that are within range of those calculated by the Clarke group (Staniforth et al., 1993) (Table 3.2).

Table 3.1. Solvent excluded sidechains in αLP core.

amino acid	N_i	α_i	n_i
Ala	24	0.5205	12.49
Cys	6	0.8238	4.943
Asp	2	1.000	2.000
Glu	4	0.6379	2.552
Phe	6	0.8704	5.223
Gly	32	0.000	0.000
His	1	0.000	0.000
Ile	8	0.9085	7.268
Lys	2	0.6496	1.299
Leu	10	0.8964	8.964
Met	2	0.9840	1.968
Asn	13	0.5238	6.809
Pro	4	0.5929	2.372
Gln	9	0.6491	5.842
Ser	20	0.4431	8.861
Thr	18	0.5127	9.228
Val	19	0.8304	15.78
Trp	2	0.9819	1.964
Tyr	4	0.8867	3.547
n total			101.0

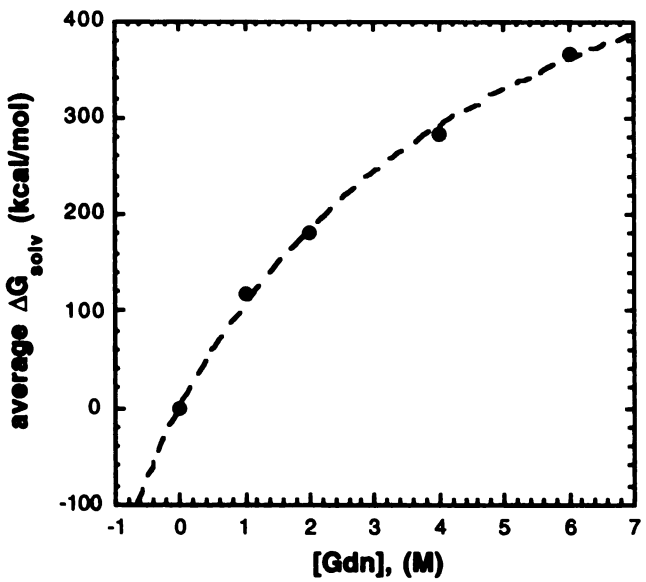


Figure 3.1 Dependence of solvation energy for an average internal sidechain in αLP on GdnHCl.

Table 3.2 Denaturation constants for internal sidechains.

protein	$\Delta G_{s,m}$ (kcal/mol)	K_{den} (M)
αLP	0.68	5.4
phosphoglycerate kinase[a]	0.47	3.1
staph nuclease[a]	0.68	4.9
"average" protein[a,b]	0.78	5.4

[a]From (Staniforth et al., 1993),Table III
[b]calculated using equations *(16)-(21)* for 55 proteins in the pdb database

Calculation of Denaturant Activity from Molarity based on hydrophobic solvation

From the variation on the solvent transfer model developed by Staniforth, Parker *et al.* (Parker et al., 1995) have proposed a method to calculate a denaturant activity corrected for nonlinearity due to the effect of denaturant on hydrophobic solvation. From fits with equation *(20)* of the free energy of transfer (ΔG_{solv}) of non-polar amino acid side chains and of the peptide group from water to specific concentrations of GdnHCl, they observe that the dependence of $\Delta G_{s,m}$ on the number of carbon atoms is linear. Therefore they attribute the non-linearity of ΔG_{solv} with GdnHCl to the $C_{0.5}$ value, which clusters fairly randomly around an average of 7.5 ± 2.1 M when plotted vs. carbon atoms. Using that value, they calculate molar denaturant activity (D):

$$D = [GdnHCl] \cdot (7.5\ M / 7.5\ M + [GdnHCl]). \qquad (22)$$

Their plots of ΔG_U vs. D for lysozyme and the N terminal domain of phosphoglycerate kinase now allow for "a more reliable extrapolation of data to a condition where D= 0."

As a next step, they propose a method for determining a temperature corrected denaturant activity from GdnHCl molarity based on the temperature dependence in the ability to solvate hydrophobic surface (Parker et al., 1998). To do this, they measured the free energy of solvation of N-Acetyl tryptophanimide (NATA), as a model for an average hydrophobic residue, in a series of GdnHCl concentrations at 15, 25, 35, and 45 °C. Using equation *(22)* (effectively treating it as a saturable binding process) they measured the change in $C_{0.5}$ and $\Delta G_{s,\,max}$ with temperature. A linear fit to the change in the $C_{0.5}$ for binding with temperature is then used to determine the $C_{0.5}$ at the experimental temperature,

$$C_{0.5}(T) = (12.3 \pm 0.2\ M) - (0.16 \pm 0.05\ M/°C)(T) \qquad (23)$$

[illegible]

which is then used to calculate the protein denaturant activity (D (T))by scaling the average value of 7.5M for K_{den} at 25°C to the temperature corrected value

$$K_{den}(T) = (7.5 / C_{0.5}(25)) \bullet C_{0.5}(T) \qquad (24)$$

for use in equation *(22)* modified as:

$$D(T) = [GdnHCl] \bullet (K_{den}(T)/(K_{den}(T) + [GdnHCl]). \qquad (25)$$

The treatment of the denaturant dependence of solvation in terms of a hyperbolic function was arrived at empirically but the fit parameters do not describe a physical model. Although this formalism and the subsequent derivations to treat temperature dependence seem to empirically work for the cases for which they were developed, the application is likely to be limited, since it is not consistent with a general physical model.

Alternatively, it is possible to treat the ΔG_{solv} of model compounds vs. GdnHCl in terms of a quadratic function, which does provide a physical model based on simple solution theories (Alonso and Dill, 1991).

$$\Delta G_{solv}(GdnHCl) = \Delta G_1[GdnHCl] + \Delta G_2[GdnHCl]^2 \qquad (26)$$

ΔG_1 accounts for the interaction of the solute with water or GdnHCl, while ΔG_2 accounts for the interaction of water and GdnHCl. Alonso *et al.* demonstrated that this functional form also provides an accurate description of the dependence of the solvation of hydrophobic amino acids on GdnHCl. In combination with statistical thermodynamic theory, a curved dependence of the free energy of unfolding of a representative protein with GdnHCl can be predicted.

The dependence of ΔG_{solv} for NATA on GdnHCl can be fit with this formalism at each temperature (Figure 3.2a) to yield ΔG_1 and ΔG_2 values with a dependence on temperature (Figure 3.2bc). Because $\Delta G_1 c$ is the free energy of binding to a protein site,

IMAGE [illegible]

and ΔG_2c^2 is the excess free energy of binding two guanidines to the protein site (Alonso and Dill, 1991), their dependencies on temperature can be fit to determine the enthalpy, entropy and heat capacity of each binding event (Table 3.3).

$$\Delta G = \Delta H(T_s) + \Delta C_p [(T-T_s) - T \ln (T/T_s)] \quad (27)$$

(T_s is the temperature at which entropy = 0)

Table 3.3 Thermodynamic parameters for GdnHCl binding at 298 K to NATA.

	ΔCp (cal/K-mol)	ΔH (cal/mol)	ΔS (cal/mol)	ΔG (cal/mol)
ΔG_1	20 ± 4	790	3.5	-253
ΔG_2	3.76 ± 0.08	105	0.4	-14.2

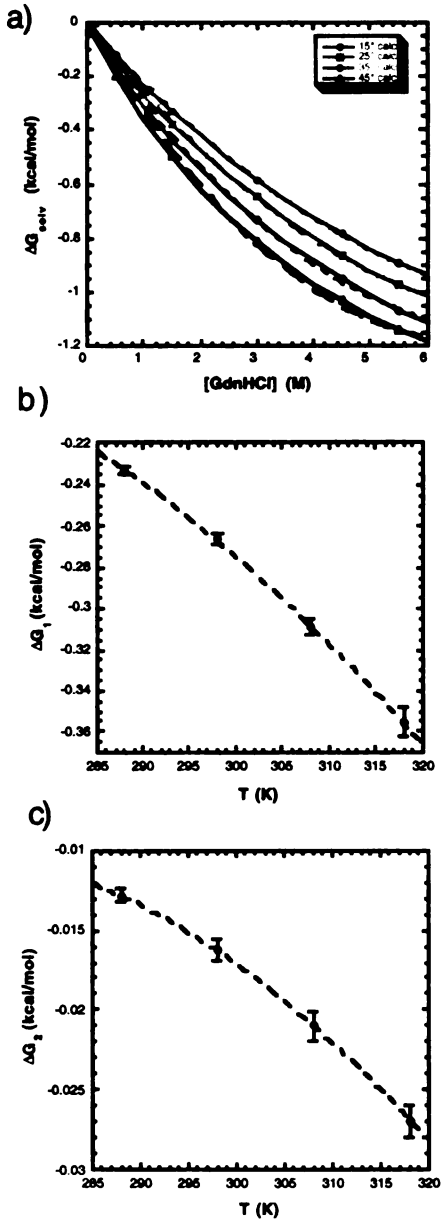


Figure 3.2. Dependence of ΔG_{solv} of NATA on GdnHCl at 15 – 45 °C. a)The solid curves were generated from the published parameters for the hyperbolic fits to the NATA ΔG_{solv} calculated based on NATA solubility in Parker *et al.* Dashed lines are the fits to equation *(26)*. b) Temperature dependence of ΔG_1 and b) ΔG_2 fit with equation *(27)*.

Eyring analysis of temperature dependence

From the temperature dependence of the rate constant for unfolding, the enthalpy, entropy and heat capacity changes associated with the unfolding transition state can be extracted (Chen and Matthews, 1994):

$$\ln(k_U/T) = A + B\,(T_0/T) + C\ln(T_0/T), \quad (28)$$

$$A = [\Delta S^{\ddagger}(T_0) - \Delta C_p^{\ddagger}]/R + \ln k_B/h, \quad (29)$$

$$B = [\Delta Cp^{\ddagger} - \Delta H^{\ddagger}(T_0)/T_0]/R, \quad (30)$$

$$C = -\Delta C_p^{\ddagger}/R. \quad (31)$$

Results and Discussion

Linear Extrapolation of Unfolding in Denaturant is not valid over a wide temperature range

Comparison of αLP unfolding by GdnHCl or urea at 4 °C indicated that the linearly extrapolated rates of unfolding in the absence of denaturant were identical within error (Figure 3.3).

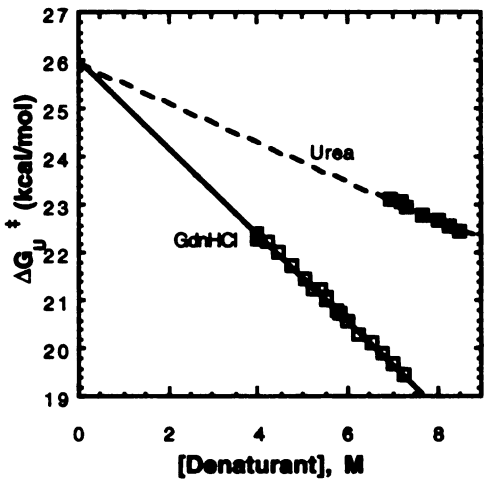


Figure 3.3 αLP unfolding rates (in terms of $\Delta G^{\ddagger}_{U}$) in GdnHCl and urea.

In addition, the ratio of either slope to the slope calculated for the equilibrium unfolding transition provided equivalent estimates of surface area exposed upon unfolding (18%) (Chapter 4). This seemed to validate a simple linear dependence of αLP's unfolding rate on either denaturant.

Measurement of the unfolding rate above 3M GdnHCl at a series of temperatures between 4 and 37 °C also appeared to support a linear dependence (Figure 3.4a). Although there does appear to be a slight decrease in m value with temperature, the slope changes only about 8% over the experimental temperature range (Figure 3.4b). An Eyring analysis of the extrapolated rates to extract the heat capacity, entropy and enthalpy of the unfolding transition state demonstrated a good fit and yielded reasonable values for those parameters (Figure 3.4c).

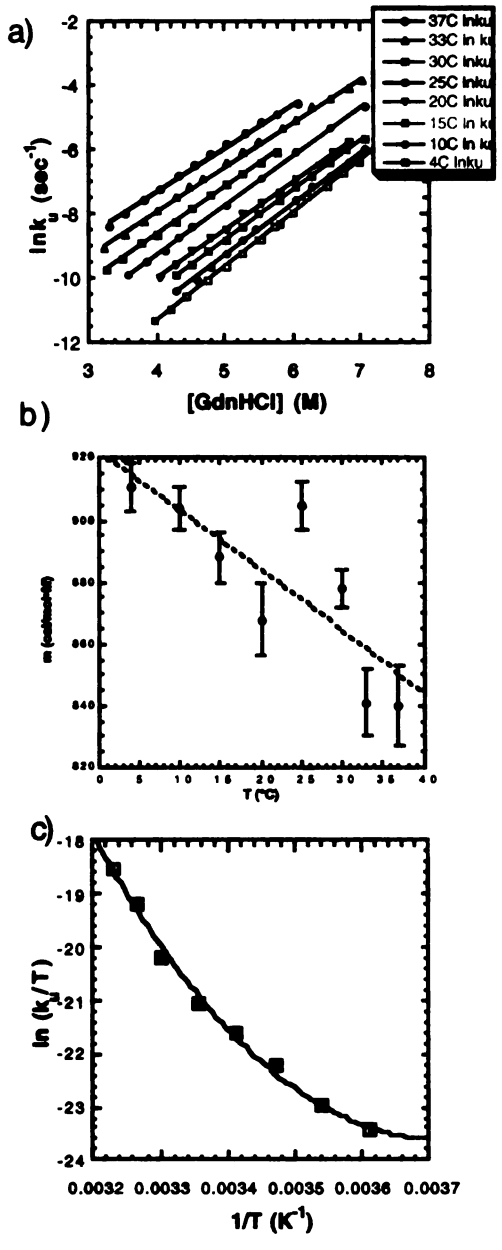


Figure 3.4. LEM applied to unfolding rates between 4 – 37 °C. a) Linear fit to ln k_u in 3 – 7 M GdnHCl. b) Variation in m values with temperature. c) Eyring analysis of linearly extrapolated unfolding rates in the absence of denaturant yields $\Delta C_p^{\ddagger} = 1.1 \pm 0.2$ kcal/mol•M, $\Delta H = 31 \pm 1$ kcal/mol, $T\Delta S = 5 \pm 1$ kcal/mol ($R = 0.9988$).

However, measurement of the unfolding rate at 50 °C, where unfolding is accelerated enough so that it is possible to measure unfolding down to very low GdnHCl

concentrations on a practical time-scale, revealed a strongly curved dependence on GdnHCl. As shown in Figure 3.5a, application of the linear extrapolation model to the data in the higher denaturant concentration range where unfolding data for most proteins is accessible leads to a poor fit to the lower concentration data. This dependence was fit extremely well by the denaturant binding model instead. Subsequent examination of lower denaturant unfolding at selected lower temperatures showed that curvature did become evident, and the data over the entire denaturant concentration range was fit much better with the denaturant binding model than the LEM (Figure 3.5 b-d).

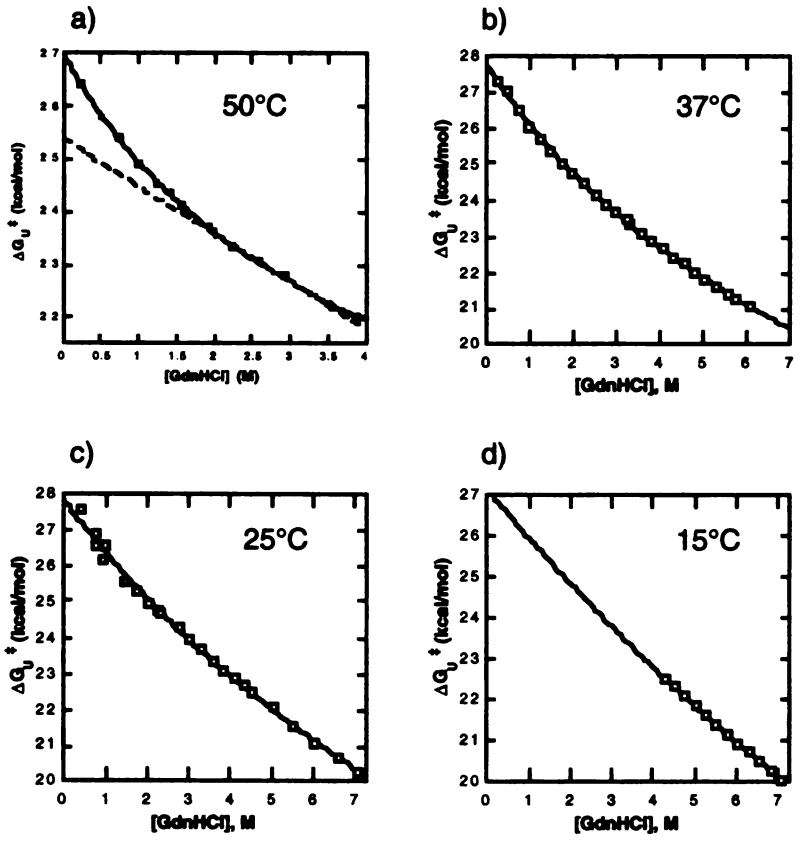


Figure 3.5. $\Delta G^{\ddagger}_{H2O}$ dependence on GdnHCl at higher temperatures.

As a test of the accuracy of the denaturant binding model extrapolation, we monitored unfolding in GdnHCl at 60 °C, where unfolding occurs fast enough to measure directly. Remarkably, the $\Delta G^{\ddagger}_{H2O}$ predicted by the denaturant binding model (26.15 ± 0.05 kcal/mol) is within error of the $\Delta G^{\ddagger}_{H2O}$ (26.20 ± 0.01 kcal/mol) calculated from the directly measured value.

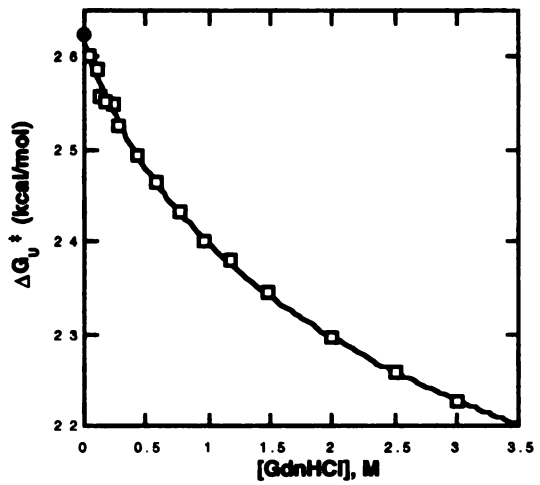


Figure 3.6. Denaturant binding fit $\Delta G^{\ddagger}_{H2O}$ vs. directly measured $\Delta G^{\ddagger}_{H2O}$ at 60 °C.

The opportunity to follow denaturant induced unfolding over such a wide temperature and concentration range for αLP illustrates the perils in assuming that a linear dependence of ΔG at high denaturant ranges validates using LEM to determine ΔG_{H2O}, and in assuming that linearity demonstrated by independent techniques at one temperature holds for another temperature. In the case of αLP, the denaturant binding model provides an accurate description of the curvature.

Temperature Dependence of Denaturant Binding Model

Although the traditional denaturant binding model ($\Delta G_{den} = \Delta G_{H2O} - \Delta nRT \ln (1+ kD)$) fits the data extremely well at all experimental temperatures where curvature is observed, close examination of the curvature with temperature reveals that it is most extreme at 60 °C, and decreases with decreasing temperature. The 15 °C data, although still well fit with the denaturant binding model, is much closer to linear, and the highly linear 4 °C data could not be fit at all with the denaturant binding model (data not shown). This changing curvature with temperature is reflected in the variation of the fit parameters Δn and k (Table 3.4). Interestingly, they display an inverse dependence on temperature, which appears to be exponential, as demonstrated by the linear dependence on temperature of their natural logarithms (Figure 3.7).

Table 3.4 Temperature dependence of denaturant binding parameters.

Temperature (°C)	Δn (binding sites)	k (binding constant)
15	30 ± 20	0.07 ± 0.03
25	20 ± 3	0.13 ± 0.03
37	11 ± 0.5	0.28 ± 0.02
50	5.6 ± 0.3	0.8 ± 0.08
60	2.9 ± 0.1	2.2 ± 0.1

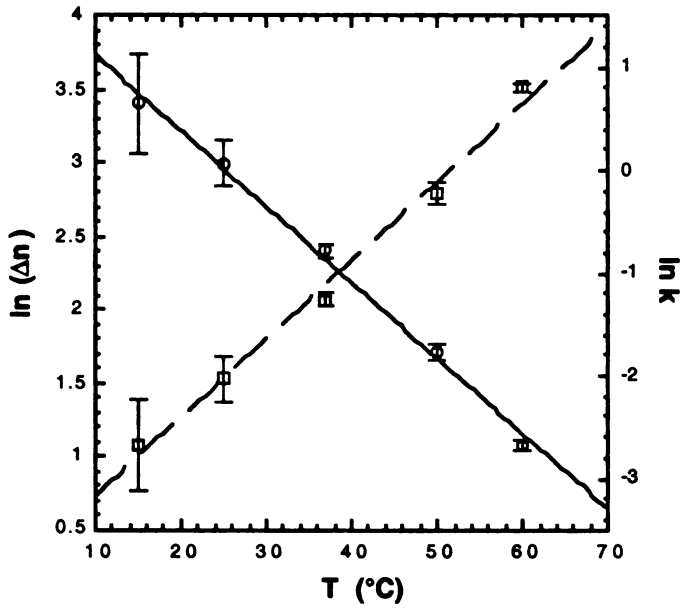


Figure 3.7 Temperature dependence of denaturant binding parameters. ln (Δn) (circles) decreases with temperature, ln k (squares) increases with temperature.

Because most other studies which have applied the denaturant binding model to unfolding data at a series of denaturant concentrations were only carried out at one temperature (Pace, 1975), an analysis of the temperature dependence of these parameters has not been previously possible. The meaning of this inverse temperature dependence is not immediately clear in terms of the mechanism of denaturant unfolding the protein. Strictly speaking it would indicate that as the temperature increases, the number of sites exposed on the unfolding transition state decreases, from 30 at 15 °C to 3 at 60 °C, while the strength of the binding increases 30-fold. The decrease in sites could indicate that the transition state is becoming less fully extended at higher temperatures, but it is hard to imagine that it is 10-fold more compact at 60 °C than at 15 °C.

Using titration calorimetry, Makhatadze and Privalov found that k for the binding of GdnHCl to three different unfolded proteins did change with temperature, but in the opposite direction, decreasing from an average of 0.8 at 10 °C to 0.52 at 40 °C (Makhatadze and Privalov, 1992). The discrepancy in the trend of the k with temperature may be due to their assumption that Δn does not change with temperature, or to a different behavior of binding a completely unfolded molecule in their case, compared to the more compact unfolding transition state in our case. The bottom line appears to be that there is a heat capacity involved in the effect of GdnHCl on unfolding, and the traditional denaturant binding model in terms of the Δn and k parameters is not adequate to accurately treat it.

However, if Δn and k are combined into an" effective m value" in analogy to the LEM m value (see Methods), by assuming standard state conditions (1 M), the m value actually corresponds to a ΔG term, which should describe the interaction of the denaturant with the protein. The variation of this term with temperature displays a clear linear dependence. With increasing temperature, the m value increases, in correspondence with the increasing curvature in the ΔG vs. denaturant with temperature (Fig 3.4.).

Remarkably, the linear fit to these m values calculated from the denaturant binding parameters at 15 – 60 °C predicts an m value at 4 °C of 890 cal/mol, which is within error of the m value of the linear fit to the 4 °C data (910 ± 20 cal/mol•M), which was confirmed by correspondence with the urea extrapolation. Therefore, we propose that this effective m value offers a measure of denaturant protein interaction in terms of energy, which quite accurately describes the degree of curvature expected in the data. It is

likely that the slope of this interaction energy with temperature will vary depending on the protein, and the degree and type of exposure of surface in the transition state or unfolded state.

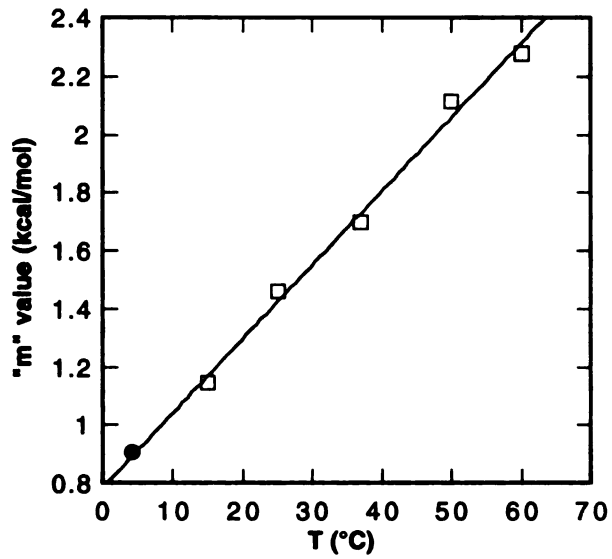


Figure 3.8 Dependence of effective "m" value on temperature. Squares represent effective "m" values calculated from denaturant binding fits. Circle denotes the measured 4 °C m value by LEM.

Temperature dependence of the denaturant binding effect is not due to the movement of the transition state

One possible source of the increased curvature in the denaturant dependence is "movement" of the transition state, i.e. it is changing conformation with denaturant and temperature. Reports of the fully denatured state changing with denaturant and temperature have also been made (Dill and Shortle, 1991). This is thought to be because the denatured state is a polymer conformation. If the monomer dislikes the solvent more in the changed conditions, the conformation will collapse more, to make more monomer-

monomer contacts. Lowering the temperature or removing denaturant therefore causes the chain to shrink. Because the transition state is compact, rather than extended, it is even more vulnerable to being affected by changing conditions.

This has been proposed to explain curvature and its temperature dependence in ln k_u with denaturant for barnase, U1A and mutants of S6 (Otzen et al., 1999). Based on monitoring denaturant and pressure dependence of unfolding kinetics, Kiefhaber and colleagues (Pappenberger et al., 2000) observe that the volume of the transition state for tendamistat does increase with GdnHCl concentration. However, they did not determine whether the transition state volume changes with temperature as well.

To test whether the increasing curvature with temperature is a property of the αLP transition state, or the GdnHCl interaction with the transition state, we monitored unfolding in urea at 60 °C, where the GdnHCl curvature is extremely pronounced. Although very slight curvature is detectable in the data, the data are fit extremely well with a linear extrapolation, leading to an m value identical to the 4 °C m value. This indicates that the transition state surface exposure is not changing at all with temperature. Therefore the changing GdnHCl curvature must be due to changes in the interactions between the protein and GdnHCl with temperature.

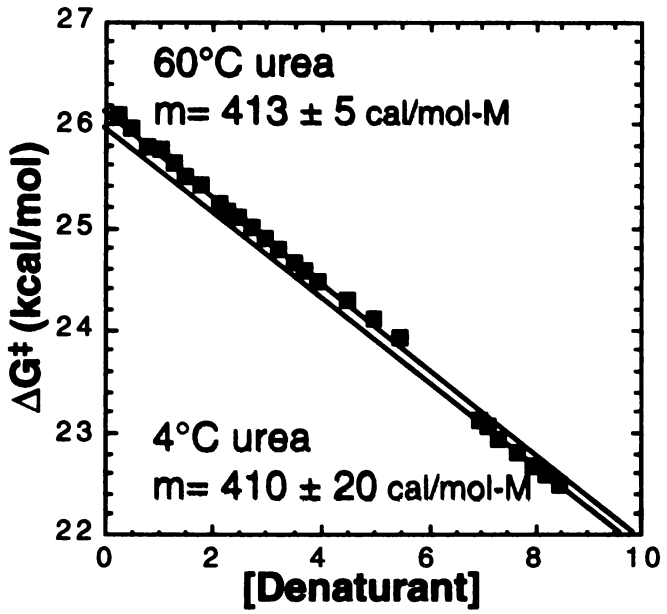


Figure 3.9 Urea denaturation at 4 and 60 °C.

Curvature is not due to ionic stabilization of the native state

One obvious possibility for the origin of the curvature is that the ionic nature of GdnHCl could be leading to a stabilization of the native state, and that stabilization increases with temperature. Stabilization of the native state by GdnHCl has been observed for ubiquitin (Makhatadze, 1999) and RNAse T1 (Mayr and Schmid, 1993). If this were the case, fitting the curvature would incorrectly extrapolate the $\Delta G^{\ddagger}_{H2O}$. However, this is ruled out both by the correspondence of the denaturant binding extrapolation with the directly measured unfolding rate at 60 °C (Figure 3.7), and the correspondence of the urea extrapolation at 60 °C, which is not ionic.

Solvent Transfer Model Fails to describe Dependence of αLP $\Delta G^{\ddagger}_{U}$ on Denaturant or Temperature

A possible source for that heat capacity is the change in the solvation of exposed hydrophobic surfaces by GdnHCl with temperature. The solubility studies of model compounds in various concentrations of GdnHCl demonstrate that dependence of the solvation free energy on GdnHCl is most curved for the transfer of the hydrophobic amino acid side chains from water to GdnHCl (Nozaki and Tanford, 1970). Obviously one would expect a temperature dependence to the solvation of the hydrophobic amino acid side chains as well.

The Clarke group has applied an alternative method for fitting curvature in the dependence of ΔG_U on denaturant based on the known solvation free energies for amino acid sidechains at 25 °C (Staniforth et al., 1993), (Parker et al., 1995). They have further developed their model to incorporate the dependence of denaturant on temperature, based on the change in solvation energy for the hydrophobic model compound NATA with temperature (Parker et al., 1998). The change in accessible surface area for each amino acid of each type buried in a specific protein core is calculated from the difference in surface area between the crystal structure and a model of an extended chain. The solvation energies can then be appropriately summed at each GdnHCl concentration, to give the change in solvation energy for an average internal side chain of that protein. For the proteins in their studies, this is adequate to describe the curvature with denaturant. Their temperature corrected denaturant activity also corrects for changing denaturant dependence with temperature (Parker et al., 1998).

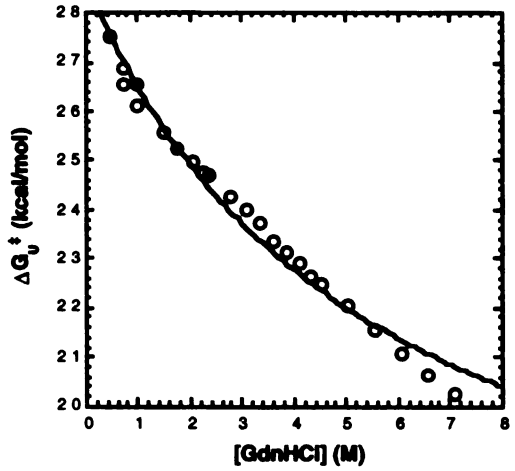


Figure 3.10 Analysis based on solvation energy for average internal sidechain in αLP.

We did this analysis for αLP at 25 °C (see Methods), but when we applied the "denaturation constants" derived from the internal solvation GdnHCl dependence to equation *(21)* ($\Delta G^{\ddagger}_{U} = \Delta G^{\ddagger}_{H2O} + n\Delta G_{sm}[D]/(K_{den} + [D])$ fitting for $\Delta G^{\ddagger}_{H2O}$ and n (number of internal side chains exposed in the transition state), the fit was quite poor (Figure 3.8). The curvature predicted by that fit is obviously more pronounced than the actual experimental data warrants.

Accordingly, when we calculated a denaturant activity using their method *(23) – (24)*, instead of becoming linear, as is the case for their proteins, a curvature in the opposite direction resulted for the $\Delta G^{\ddagger}_{U}$ dependence on denaturant at 25 °C (Figure 3.11c). This type of over-correction of curvature in the denaturant dependence using the Clarke denaturant activity calculation has also been seen by Oliveberg and colleagues (Otzen et al., 1999).

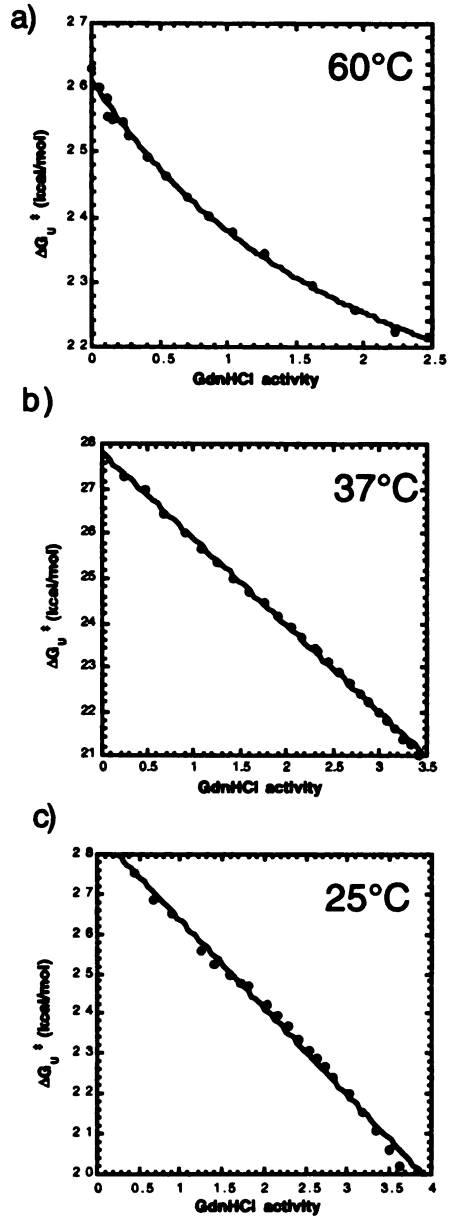


Figure 3.11 $\Delta G^{\ddagger}_{U}$ vs. GdnHCl activity calculated using Parker *et al.* (Parker et al., 1998)

In addition, correction to denaturant activity at higher temperatures resulted in a shifting curvature from the strongly opposite curvature at lower temperatures (Fig.

3.11bc) the pronounced upward curvature (as in the original $\Delta G^{\ddagger}_{U}$ vs. GdnHCl M data) at 60°C (Fig. 3.11a).

This could be a result of the αLP transition state exposing surface with an opposing effect on solvation energy (i.e. polar surface interacting with GdnHCl due to its ionic nature), whereas unfolding their proteins doesn't expose such surface. Another possibility is that the model for analyzing the transfer data for sidechains only to determine the solvation energy for unfolding in GdnHCl is over-predicting the curvature, or it could be revealing the limitations of the non-physical hyperbolic treatment of the solvation energies of the protein.

Physically based Quadratic Formalism

As described in Methods, a more physically realistic model for the effects of GdnHCl on hydrophobic solvation uses a quadratic formalism (Alonso and Dill, 1991). If we plot ΔG_U vs. ΔG_{solv} for NATA at the appropriate temperature and denaturant concentration calculated with the quadratic formalism, then the effects of hydrophobic solvation should be accurately removed. This does decrease the curvature without causing downward curvature at lower temperatures, but significant curvature still remains (Figure 3.12).

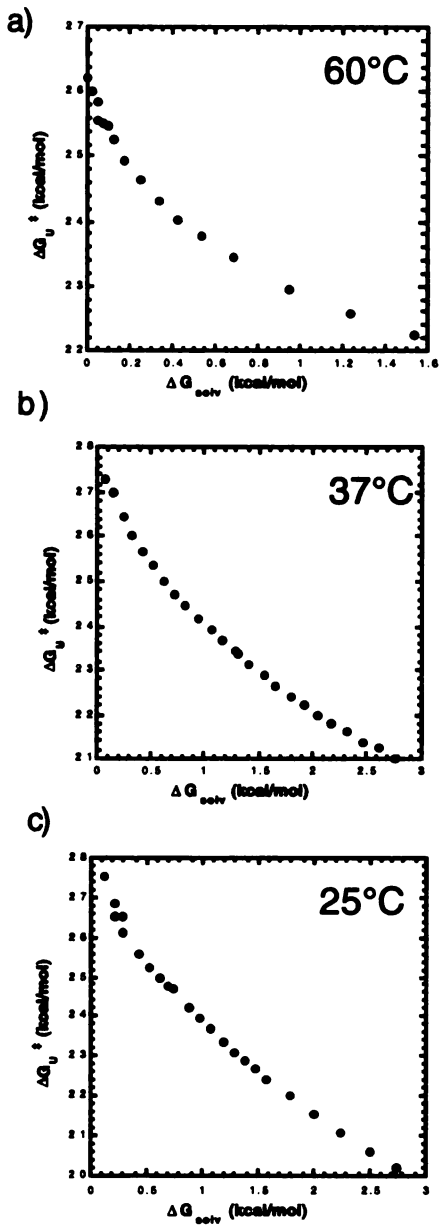


Figure 3.12 $\Delta G^{\ddagger}_{U}$ vs. GdnHCl corrected for hydrophobic solvation energy (ΔG_{solv}).

This indicates that, contrary to common assumption (Alonso and Dill, 1991), (Makhatadze, 1999), (Parker et al., 1995), the curvature in ΔG_U and $\Delta G^{\ddagger}_{U}$ dependence on GdnHCl does not solely result from the hydrophobic solvation effects. Although for

some proteins, correction for that does lead to linearity, in αLP and presumably other proteins there are other effects that can lead to curvature. For αLP, which is highly charged (pI = 10) it is very possible that interactions of polar surfaces with the ionic GdnHCl molecule are involved.

Conclusions

The results of the characterization of the unfolding kinetics of αLP over an extensive temperature and denaturant range lead to important cautions pertaining to the analysis of GdnHCl induced unfolding data. By following the unfolding kinetics from the high GdnHCl concentrations typically accessible through to zero denaturant, we conclusively demonstrated that apparent linearity in the dependence of ln k_u / $\Delta G^{\ddagger}_U$ on GdnHCl concentration at those high concentrations can belie significant curvature which is revealed only at lower concentrations. Therefore extrapolations from apparently linear high concentrations can lead to serious errors in thermodynamic parameters calculated based on them. For αLP, an Eyring analysis based on the linear extrapolations from measurements between 3 – 7 M GdnHCl was incorrect compared to the correct fit to the curved data between 0 – 7 M GdnHCl (Chapter 4) by +13 kcal/mol for $\Delta H^{\ddagger}$ and + 14.5 kcal/mol for $T\Delta S^{\ddagger}$. This underscores the importance of confirming extrapolations from high denaturant concentrations through coincidence of different denaturants, or measurement by other means such as DSC, or HX under low to zero denaturant concentration.

A second important caution based on these results is that there can be a distinct temperature dependence to the GdnHCl-protein interaction. In the case of αLP, this leads

to increasing curvature in the dependence of ln k_u and $\Delta G^{\ddagger}_U$ on GdnHCl concentration with increasing temperature. As a consequence, extrapolations need to be confirmed not only at one temperature, but over the entire experimental temperature range.

We demonstrated that the denaturant binding model provides an extremely accurate fit of the curvature in $\Delta G^{\ddagger}_U$ with GdnHCl at all temperatures for αLP. At 60 °C it predicts a $\Delta G^{\ddagger}_{H2O}$ within error of the $\Delta G^{\ddagger}_{H2O}$ based on the directly measured k_u. The very slight curvature observed at 60 °C in urea suggests that the LEM model is actually lacking adequate sophistication to model urea unfolding as well. For urea however, the difference in $\Delta G^{\ddagger}_{H2O}$ extrapolated by linear extrapolation vs. denaturant binding is insignificant: both are within error of the $\Delta G^{\ddagger}_{H2O}$ based on the directly measured k_u. Therefore, we suggest using urea when possible to monitor unfolding, and using the denaturant binding model to treat GdnHCl dependent unfolding data.

However, parameters describing the denaturant dependence (Δn and k in the denaturant binding model analysis) cannot be assumed to be independent of temperature, as is sometimes done. For αLP, the number of binding sites (Δn) and the GdnHCl binding constant (k) display an inverse dependence on temperature which is not readily interpretable. We propose combining them into a single "effective m value", with units of kcal/mol under standard state conditions (1 M), corresponding to an energy of denaturant interaction.

The "effective m value" calculated from the denaturant binding fits varies linearly with temperature, describing an increasing energy of denaturant interaction with temperature, corresponding to the increasing curvature observed. At 4 °C, the "effective m value" calculated based on the linear fit is identical to the measured m value using

LEM, which was confirmed by urea extrapolation as well. Therefore this "effective m value" may provide a useful means for describing the energy of denaturant interaction with a protein, with a slope with temperature that will vary depending on the protein, and can be used for comparison.

The curvature we observe in dependence of $\Delta G^{\ddagger}_{U}$ on temperature is in part due to the changing ability of GdnHCl to solvate hydrophobics with temperature. The hydrophobic solvation contribution to the denaturant effect can be removed by plotting the $\Delta G^{\ddagger}_{U}$ vs. a temperature appropriate hydrophobic solvation energy calculated using a quadratic formalism to describe the change in NATA solvation by GdnHCl with temperature. Despite lessening the extent of curvature, it nevertheless remains and changes with temperature. (The Parker method of calculating a denaturant "activity" based on the NATA solvation data is not valid for αLP: curvature remains at higher temperatures, and switches to a downward curvature at lower temperatures.) Therefore, contrary to common assumption, the change in hydrophobic solvation with temperature is not the sole origin of curvature in GdnHCl dependence.

Another possibility for changing curvature in the dependence of $\ln k_u / \Delta G^{\ddagger}_{U}$ on denaturant and temperature is movement of the transition state with changing conditions. It may be that the αLP transition state does change some with GdnHCl concentration: Kiefhaber and colleagues recently showed convincingly that the tendamistat transition state increases in volume with GdnHCl (Pappenberger et al., 2000). However, we showed that the αLP transition state is not changing with temperature. The m value based on LEM determined by unfolding in urea is exactly the same at 4 and 60 °C,

indicating that exposure of hydrophobic surface in the transition state is the same at both temperatures.

Therefore the change in curvature in GdnHCl dependence of $\Delta G^{\ddagger}_{U}$ with temperature must be a result of a change in the denaturant- protein interaction in addition to the change in hydrophobic solvation. We suggest that electrostatic interactions between the ionic GdnHCl molecule and the transition state may be a cause of curvature. It is likely that electrostatics will contribute to curvature in GdnHCl dependence of $\Delta G^{\ddagger}$ for highly charged proteins like αLP (pI = 10), but may not be a factor in more neutral proteins. Therefore some proteins can display a linear dependence of ΔG or $\Delta G^{\ddagger}_{U}$ on GdnHCl once the hydrophobic component is removed (Parker et al., 1998).

In addition, there may be more of a propensity for curvature in $\Delta G^{\ddagger}_{U}$ on GdnHCl than ΔG, depending on whether the Gdm^{+} or Cl^{-} ions are screening repulsive or attractive charges close in proximity in the transition state, compared to the fully extended unfolded state. Electrostatic interactions may also contribute to curvature in an opposite direction, providing a reason why some proteins display linearity in ΔG_{U} with GdnHCl concentration, even though there should be curvature in unfolding all proteins due to hydrophobic solvation.

Future Directions

General formalism incorporating curvature in dependence of ΔG_{U} and $\Delta G^{\ddagger}_{U}$ on GdnHCl and temperature

The preceding results would suggest that the solution to non-linearity possibilities would be to use urea as the unfolding agent. However, urea has the disadvantage of

decomposing over time ($t_{1/2}$ = 8 days (Pace, 1986)) and at high temperatures to form cyanate, which can affect the denaturing ability of the urea as well as may interact with the protein. In addition, urea is 2 – 3 times weaker than GdnHCl (Pace, 1986) and many proteins of interest are very stable and only unfolded in high concentrations of GdnHCl. Also, as seen in Figure 3.9, there is a slight curvature to the urea data at 60 °C as well, suggesting that although a linear extrapolation is very close, a more accurate treatment would incorporate curvature.

Since a traditional denaturant binding model does fit the GdnHCl data extremely well, but leads to a non physical dependence of the number of sites exposed on temperature, we are developing a modified denaturant binding formalism, which treats the temperature dependence solely in terms of the binding. The binding is treated with the quadratic formalism based in solution theory, allowing a physical interpretation of the parameters and their temperature dependence. Using the temperature dependence of the binding parameters for NATA, we will be able to incorporate the changing contributions of solvating hydrophobic residues by GdnHCl with temperature. This will allow us to derive an expression which inputs those known parameters, and fits the measured change in ΔG_U and $\Delta G^{\ddagger}_U$ with temperature and denaturant, yielding the ΔG_{H2O} and $\Delta G^{\ddagger}_{H2O}$ and parameters describing the change in ΔG not due to the hydrophobic effect. Although there may be more solvent contributions to the ΔG dependence than purely hydrophobic ones, this will at least provide a means for removing the known temperature dependence of hydrophobic solvation in the analysis of denaturant unfolding data.

Electrostatic contributions to temperature dependent change in denaturant effect

Figure 3.12 showed that the removal of the temperature dependence of GdnHCl hydrophobic solvation (as modeled by NATA transfer) still results in curvature in the ΔG vs. GdnHCl, and that curvature still has a temperature dependence. This could be due to the fact that GdnHCl is ionic, and αLP is highly charged. An obvious experimental approach to determine whether electrostatics are the origin of the remaining curvature in αLP would be to look at the effects of salt and pH on unfolding in urea.

In analogy to the NATA transfer experiments, it would be useful to determine the temperature dependence of the transfer of models for + and - charged amino acids into GdnHCl. This would allow the removal of the effect of GdnHCl on polar solvation in addition to the hydrophobic solvation effect on the denaturant dependence of unfolding. In terms of developing a formalism for analyzing denaturant dependence of unfolding which allows the incorporation of parameters completely describing the general effects of GdnHCl on solvation alone, determining the temperature dependence for the transfer of each individual amino acid would be the most accurate.

Chapter 4. Thermodynamic Nature of Kinetic Stability in αLP.

Introduction

αLP is synthesized as a precursor molecule, which is processed to the mature 198aa active protease by autolytic cleavage of the 166aa pro region (Silen et al., 1988), (Silen et al., 1989). Folding of αLP requires the activity of this pro sequence, which appears to function by reducing a large kinetic barrier that prevents an inactive molten globule-like intermediate from reaching equilibrium with the active native conformation. We have recently demonstrated that after cleavage and proteolytic degradation of the pro region, the native state is actually kinetically trapped in a conformation with less thermodynamic stability than the both the intermediate and the completely unfolded state (Sohl et al., 1998). This highly unusual result suggests that the evolution of kinetic barriers (together with catalysts like the pro region) as a means to gain stability can render thermodynamic stability unnecessary.

Based on titration calorimetric studies, we determined that the native state is favored over the intermediate by 18 kcal/mol, implicating the entropy change as the cause of the unusual balance of stability favoring the unfolded states. A possible source for excess favorable entropy in unfolding αLP was revealed from examining the sequence of αLP, which has 16% glycines, compared to 7% on average in protein sequence (Creighton, 1993) and 9% in chymotrypsin, a thermodynamically stable homologue. Because glycines lack a sidechain, unfolded states can sample more conformations, resulting in ~0.7 kcal/mol contribution to the entropy favoring unfolding (D'Aquino et al., 1996). In addition, we proposed that by allowing tighter turns and better packing in the folded state, a large number of glycines could lead to a more rigid native conformation,

possessing smaller entropy than usual. Indeed, there is evidence that the αLP native state is highly rigid by a number of measures of conformational flexibility (protease sensitivity, B factors, HX protection factors).

Here we investigate further the thermodynamics and surface accessibility of the components of the αLP folding pathway, in particular the kinetic barrier. We determine surface area burial as a measure of the degree of folding of the intermediate and transition states. From the dependence of the unfolding kinetics on temperature, we determine the activation parameters of the unfolding barrier and how they change with temperature. This combined insight into the balance of entropic and enthalpic penalties and the degree of surface area exposure in the transition state lead to a model for the physical properties underlying the kinetic stability of αLP. In addition, the combination of the surface accessibility and the unfolding activation parameters with the previous thermodynamic results allows us to construct a nearly complete folding reaction profile for αLP.

Probing the thermodynamics of αLP 's high barrier will provide insight not only into the αLP landscape, but into kinetic barriers in general. Kinetic barriers separating two biologically relevant conformations are evident throughout nature (Baker and Agard, 1994). Serpins rearrange from an inhibitory metastable conformation to an inactive stable one, at a physiologically important rate determined by a high but not insurmountable barrier between them (Baker and Agard, 1994). Other systems may have evolved barriers which are higher and therefore require additional assistance such as the action of a pro region in folding many extracellular proteases in addition to αLP (Baker et al., 1993), (Eder and Fersht, 1995). Amyloidogenic proteins appear to populate two different conformations, a normal cellular one and a misfolded fibril forming one (Kelly,

1996). In these cases the ability to seed fibril formation of normal protein through adding a small amount of already formed fibril indicates that the seed is able to lower the barrier between the two conformations to catalyze fibril formation. Although the barrier heights and relative stabilities of the separated conformations will be variable depending on the protein, investigation of the basis for such high barriers may provide insight into how to manipulate them, and how to engineer greater stability through barriers in target proteins.

Methods

Materials

GdnHCl and Urea (ultrapure) were from ICN (Cleveland, OH, U.S.A.) Stocks were filtered through a 0.22 μM filter and the concentration was determined using refractive index measurements (Pace 86). All other chemicals were from Fisher.

SA195 αLP expression

D1210 *E. coli* transformed with the bicistronic pro region SA195 construct (Sohl ref)were grown at 37 °C in 4.5 – 6 L LB media or LXY media (15 g yeast extract, 10 g NaCl / liter) buffered at pH 6.8 with 60 mM ACES to an OD_{600} of 1.8 (LB) or 2.5 (LXY), shifted to 12 °C for two hours, then induced with 0.1 mM IPTG.

SA195 aLP purification

Supernatant from the cells harvested 6 – 7 days post induction was diluted 4 – 5-fold with dH_2O and brought to pH 4.5 – 5. 50 mls S-Seph beads were added and mixed overnight. Beads were collected and washed with 10 volumes 10 mM NaOAc pH 5.0, then 5 volumes either 10 mM Glycine, pH 9.6 or 10 mM Hepes, pH 8.4. SA195 was eluted in a noncovalent complex with the pro region using 10 mM Glycine pH 9.6, 200

mM NaCl. Incubation with trypsin coupled beads at pH 7.5 or 0.1 mg/ml pepsin at pH 3 for 1 – 12 hours followed by dialysis against or 1:4 dilution with 10 mM NaOAc pH 5 digested the pro region. Post protease treated material was loaded onto either Pharmacia MonoSHR10/10 HPLC and eluted with 10 mM NaH_2PO_4 10 – 500 mM NaOAC, pH 7.2, or Vydac VHP81010 and eluted with 10 – 500 mM NaOAc pH 5. Final protein was filtered through a 0.22 μM filter and judged pure as determined by silver stained gel, mass spectrometry and N-terminal sequencing. The x-ray structure of SA195 is virtually identical to that of the wild type (S.S.J., S.D. Rader and D.A.A. unpublished data).

Calculation of Buried Surface Area

The program ACCESS was used to determine accessible surface area, using a probe radius of 1.4 Å, slice width 0.25 Å, and atomic radii by (Richards, 1977). For native αLP, the pdb file 2alp was used. An extended model of the unfolded molecule was built using InsightII as described in Myers *et al.* (Myers et al., 1995) O, N, and S atoms were counted polar, C was counted nonpolar.

Calculation of m_{N-U}

ΔASA_{N-U} was corrected for the three disulfides in αLP (-900 $Å^2$/ disulfide) (Myers) and used in the equations:

$$m_{N-U}, (GdnHCl) = 953 + 0.23(\Delta ASA)$$

$$m_{N-U}, (urea) = 368 + 0.11(\Delta ASA)$$

Kinetic Unfolding Experiments

Unfolding reactions were done in 10 mM KOAc pH 5.0 and initiated by manual mixing to a final SA195 concentration of 0.1 to 1.75 μM for fluorescence or 7 μM for circular dichroism studies. Guanidine denaturation was carried out at 4, 15, 25, 37 50,

60, 70 °C. Refractive index measurements were used to accurately measure the concentration of GdnHCl in each sample (Pace, 1986). Fluorescence experiments were done using an 8100SLM-Aminco Fluorimeter connected to an external thermostat bath. Excitation was at 283 nm and emission at 322 nm. Circular Dichroism experiments were monitored at 225 nm using a Jasco J-715 Spectropolarimeter with a Peltier temperature control. Nonlinear regression analysis with the program Kaleidagraph (Abelbeck Software, Reading, PA) was used to obtain the monoexponetial rate constants.

Analysis of Unfolding Kinetics

Applying transition state theory by the method of Chen, *et al.* (Chen and Matthews, 1994) the activation free energies at each denaturant concentration were determined from the rate constants:

$$\Delta G^{\ddagger}_{U} = -RT \ln k_u h / k_B T. \qquad (1)$$

Below 25 °C the linear extrapolation method (LEM):

$$\Delta G^{\ddagger}_{U} = \Delta G^{\ddagger}_{H2O} - m\,[den] \text{ (Pace, 1986)} \qquad (2)$$

was used to determine the activation free energy of unfolding in the absence of denaturant. At 25 °C and above (see Chapter 3), the data were best fit with the denaturant binding model, which assumes a discrete number of binding sites for denaturant on the unfolded molecule:

$$\Delta G^{\ddagger}_{U} = \Delta G^{\ddagger}_{H2O} - \Delta n\, RT \ln(1 + k[den]), \qquad (3)$$

where Δn= difference in # binding sites between the native and transition state ensemble (normally the unfolded molecule for equilibrium unfolding) and k = equilibrium binding constant for denaturant (Pace, 1986). At 60 and 70 °C the unfolding rate was measured directly in the absence of denaturant.

Analysis of temperature dependence

In itself, the $\Delta G^{\ddagger}_{U}$ indicates the height of the barrier, but to understand the high energy of the barrier in terms of possible physical contributions, it is necessary to probe the components underlying the free energy: enthalpy and entropy.

$$\Delta G^{\ddagger}_{U} = \Delta H^{\ddagger}_{U} - T\Delta S^{\ddagger}_{U}. \qquad (4)$$

Because both $\Delta H^{\ddagger}_{U}$ and $\Delta S^{\ddagger}_{U}$ change with temperature as determined by the heat capacity of the transition state,

$$\Delta H^{\ddagger}_{U} = \Delta H_0 + \Delta C_p{}^{\ddagger}_{U}(T - T_0) \qquad (5)$$

$$\Delta S^{\ddagger}_{U} = \Delta S_0 + \Delta C_p{}^{\ddagger}_{U} \ln (T / T_0) \qquad (6)$$

monitoring the change in ΔG with temperature allows the extraction of $\Delta H^{\ddagger}_{U}$, $\Delta S^{\ddagger}_{U}$ and $\Delta C_p{}^{\ddagger}_{U}$ using an equation derived from equations *(4) – (5)* (Chen and Matthews, 1994):

$$\ln(k_u/T) = A + B\,(T_0/T) + C \ln(T_0/T), \qquad (7)$$

$$A = [\Delta S^{\ddagger}_{U}(T_0) - \Delta C_p{}^{\ddagger}_{U}]/R + \ln k_B / h,$$

$$B = [\Delta C_p{}^{\ddagger}_{U} - \Delta H^{\ddagger}_{U}(T_0)/T_0]/R,$$

$$C = -\Delta C_p{}^{\ddagger}_{U} / R.$$

Results

Solvent Accessibility of Intermediate and Transition States

The exposure of solvent accessible surface area provides a measure of how folded a component of a folding pathway is, relative to the fully folded state (Jackson, 1998). Obviously it is not possible to obtain a high resolution structure of the molten globule intermediate and transition states of αLP in order to directly calculate the solvent accessible surface area. However, the slope of the denaturant dependence for unfolding (m-value) has been shown to be proportional to the amount of surface exposed on

unfolding (Myers et al., 1995). Therefore the solvent accessibility in each state relative to the native state can be determined based on the ratio of the m-value for the appropriate transition (m_{N-TS} or m_{I-U}) and the m-value for complete unfolding of the native state (m_{N-U}). These are used to determine a parameter known as β, describing the position of conformations along the reaction coordinate for folding (Jackson, 1998) or fraction folded:

$$\beta_{TS} = 1 - m_{N-TS} / m_{N-U} \qquad \beta_I = m_{I-U} / m_{N-U}$$

with a β value of 1 indicating burial of surface equivalent to the native state, and β value of 0 indicating as much surface exposure as the completely unfolded molecule.

The large kinetic barrier separating the native state from the unfolded conformations prevents the measurement of equilibrium unfolding to directly determine m_{N-U},. However it is easily calculated from the known ΔASA_{N-U} based on the correlation between ΔASA_{N-U} and m_{N-U} (Myers et al., 1995) (see Methods).

The β parameters were determined from the calculated m_{N-U} values and the measured m_{N-TS} and m_{I-U} from previous experiments monitoring the urea induced unfolding of the intermediate (Sohl et al., 1998) and the GdnHCl and urea induced unfolding of the native to the transition state (Figure 3.3) (Table 4.1). The intermediate buries about half of the surface buried in the native state, while the transition state is highly native like, with 80% of the surface buried.

Both of these results are in keeping with other experimental evidence. The intermediate is known to be a molten globule, and by gel filtration elutes at a molecular weight about twice that of the native state (Baker et al., 1992b), so the result that it is only about half folded based on surface area makes sense. A highly native-like transition

state was expected, based on the fact that the αLP pro region binds very tightly to it as well as the native state, indicating that they must share structural features (Baker et al., 1992a), (Sohl et al., 1998).

Table 4.1 M value analysis to determine β for I and TS.

Parameter	Transition	measured value	N-U (calculated*)	β (fxn folded)
urea m value (cal/mol•M)	I - U	1100 ± 60	2250	.49
urea m value (cal/mol•M)	N - TS	400 ± 20	2250	.82
GdnHCl m value (cal/mol•M)	N - TS	910 ± 10	4900	.81

Activation parameters for unfolding

Although the transition state exposes less than 20% of the surface area buried in the native state, unfolding to it from the native state is accompanied by a loss in free energy of 26 kcal/mol (Sohl et al., 1998). In order to understand the basis for this huge barrier, we probed the enthalpic and entropic components of the barrier, through measuring change in the unfolding rate constant with temperature.

Unfolding monitored by loss of tryptophan fluorescence or elliptical absorbance at 222 nm by circular dichroism results in identical single exponential rate constants (Figure 3.2). As these techniques provide measures of tertiary and secondary structure, respectively, this indicates that αLP unfolds in a single cooperative two-state transition. Therefore we could use minimal amounts of protein to perform an extensive series of

fluorescence unfolding experiments at temperatures ranging from 4 – 70 °C and GdnHCl concentrations between 0 – 7 M.

The linear extrapolation of ln k_u vs. GdnHCl to obtain the unfolding rate in the absence of denaturant at 4 °C was previously published (Sohl et al., 1998). Due to a temperature dependent denaturant binding effect, data at 25 °C and above were best fit with a denaturant binding model to determine the unfolding rate in water (effect described and fits shown in Chapter 3). At 60 and 70 °C, unfolding could be measured directly in the absence of GdnHCl (Figure 4.1).

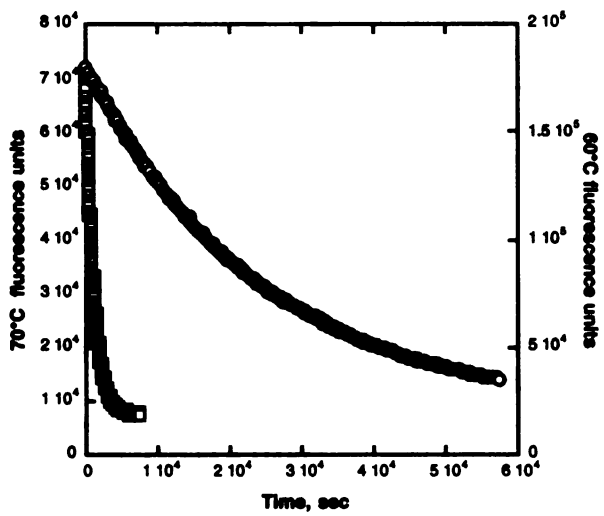


Figure 4.1 Unfolding of αLP in the absence of denaturant monitored by fluorescence at 60 and 70 °C. 60 °C unfolding in green circles (k_u = 4.122 x 10^{-5} ± 8 x 10^{-8} sec^{-1}), 70 °C unfolding in blue squares (k_u = .000924 ± 1 x 10^{-6} sec^{-1})

Using the rates of unfolding in the absence of GdnHCl, an Eyring analysis was performed to fit the dependence of the unfolding rate on temperatures and extract the thermodynamic activation parameters (Chen and Matthews, 1994) (Figure 4.2).

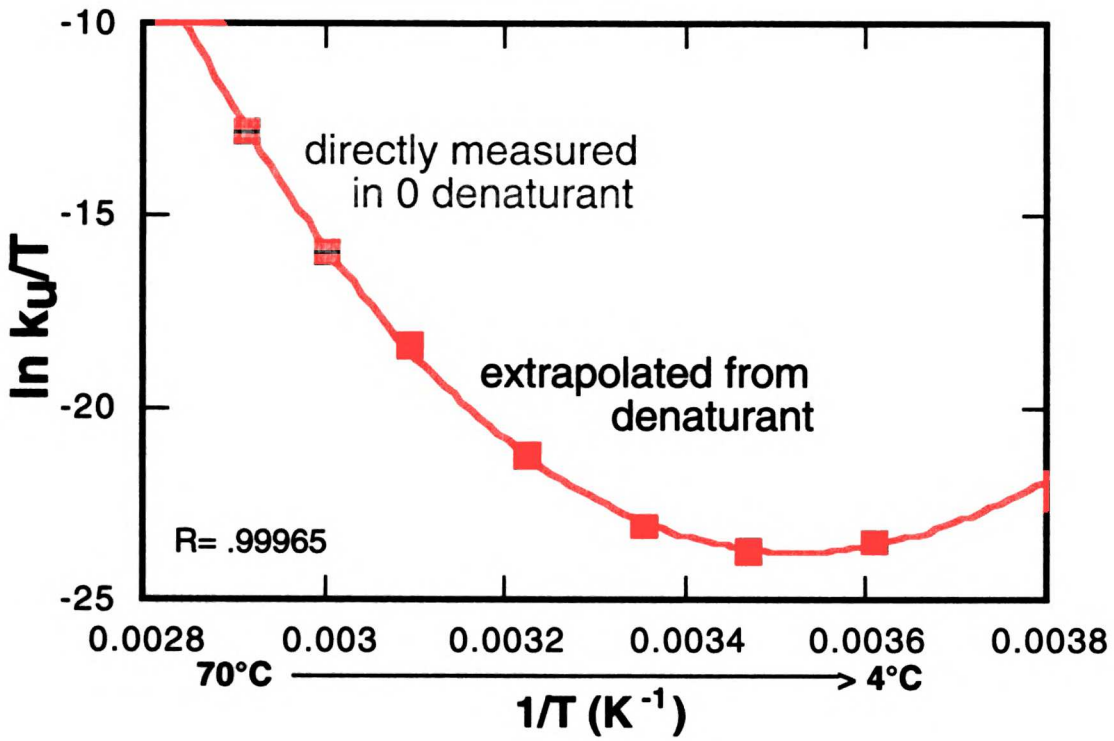


Figure 4.2 Eyring analysis of unfolding barrier. ln (k_u/T) (black: k_u extrapolated as described from denaturant, blue: directly measured in zero denaturant) vs. 1/T between 4 – 70 °C is well fit by an Eyring analysis (R = 0.99973). At 25 °C, $\Delta C_p^{\ddagger}{}_U = 1.305 \pm 0.04$, $\Delta H^{\ddagger}{}_U = 18.1 \pm 0.6$, and $\Delta S^{\ddagger}{}_U = -0.032 \pm 0.002$.

As indicated by the significant curvature with temperature, a substantial heat capacity difference (1.3 kcal/mol•K) exists between the native and transition states. This points to a large exposure of hydrophobic surface area in the transition state (Myers et al., 1995), (Makhatadze and Privalov, 1995), corresponding to about 40% of the hydrophobic surface area exposed on total unfolding (based on a predicted $\Delta C_{p,N-U}$ value of 3.3

kcal/mol•K for αLP's known ΔASA_{N-U} (Table 4.1)). This is 20% more surface exposure than predicted by m-value analysis. Such a discrepancy between $\Delta C_{p,N-U}$ and m-value predictions has been noted before (Plaxco et al., 1998a), (Plaxco et al., 1998b), (Baskakov and Bolen, 1999) and the origins are uncertain. However, as mentioned above, the ability of the transition state to bind with subnanomolar affinity to the pro region (Sohl et al., 1998), which is also a subnanomolar inhibitor of the native enzyme (Baker et al., 1992a) strongly suggests that the transition state is 82% native-like, rather than only 60%.

Because of the large heat capacity, the relative contributions of enthalpy and entropy to the unfolding barrier change significantly over a physiologically relevant temperature range (Figure 4.3). At low temperatures, the barrier is entirely entropic in origin, but as the temperature increases, the enthalpic component begins to dominate. The unfavorable entropy at lower temperatures indicates that the entropic penalty for ordering solvent around the newly exposed hydrophobic surface area is greater than the increase in configurational entropy on unfolding. However, at higher temperatures, the favorable configurational entropy dominates over the unfavorable solvation entropy, leading to an overall favorable entropic term. The large enthalpic penalty faced for unfolding at high temperatures arises from the loss of favorable native state interactions. The favorable enthalpy at lower temperatures indicates that the unfavorable loss of interactions must be outweighed by favorable interactions between the solvent and the exposed surface of the transition state.

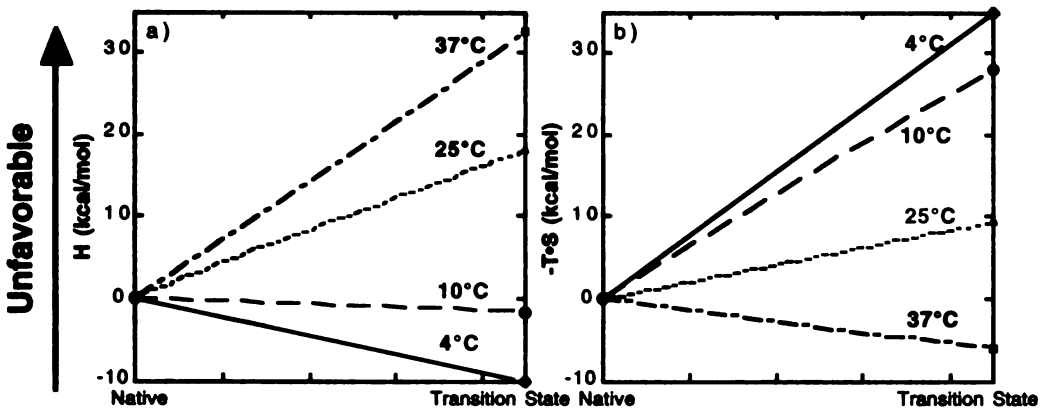


Figure 4.3 Change in unfolding barrier with temperature. a) Enthalpy becomes less favorable as the temperature increases. b) Entropy becomes more favorable as the temperature increases.

Activation parameters for folding

The origin of the refolding barrier, which is even larger than the unfolding barrier, is also of great interest. Unfortunately, because of the extreme sensitivity of the intermediate to temperature, it is not possible to measure refolding over a temperature range large enough to accurately determine the temperature dependence of the entropy and enthalpy of refolding. However, ΔH_{N-I} and ΔS_{N-I} are known at 10 °C (Sohl et al., 1998), and therefore it is possible to subtract the unfolding activation parameters at 10 °C to determine the contributions of enthalpy and entropy to the refolding barrier at 10 °C (Table 4.2).

Table 4.2. Activation parameters for the unfolding barrier and refolding barrier.

Parameter	N - TS	I - TS	N - I
ΔG (kcal/mol)	26 ± 1	30 ± 1	- 4 ± 0.1*
ΔH (kcal/mol)	- 2 ± 1	-20 ± 2	18 ± 2
TΔS (cal/mol)	- 28 ± 1	-50 ± 2	22 ± 2

* measurement at 4°C

Based on this calculation, the refolding barrier at 10 °C appears entirely entropic, while the enthalpy term is actually favorable for folding. This can be understood based on the fact that the intermediate is a molten globule ensemble, stabilized entirely by entropy, and in fact has more than usual entropy, due to the high number of glycines (Sohl et al., 1998). Therefore folding to the highly native-like transition state would result in a great loss of configurational entropy, but a gain in enthalpy from forming some favorable interactions in the transition state.

Discussion

Reaction Profile for αLP folding

The combined thermodynamic information for the folding of αLP at 10 °C is summarized in terms of a folding reaction coordinate, with the position of each state specified by the β parameter (Figure 4.4). The most striking feature of the folding reaction profile is that both the folding and unfolding barriers are entirely entropic in origin. The enthalpy, however, is favorable for both folding and unfolding

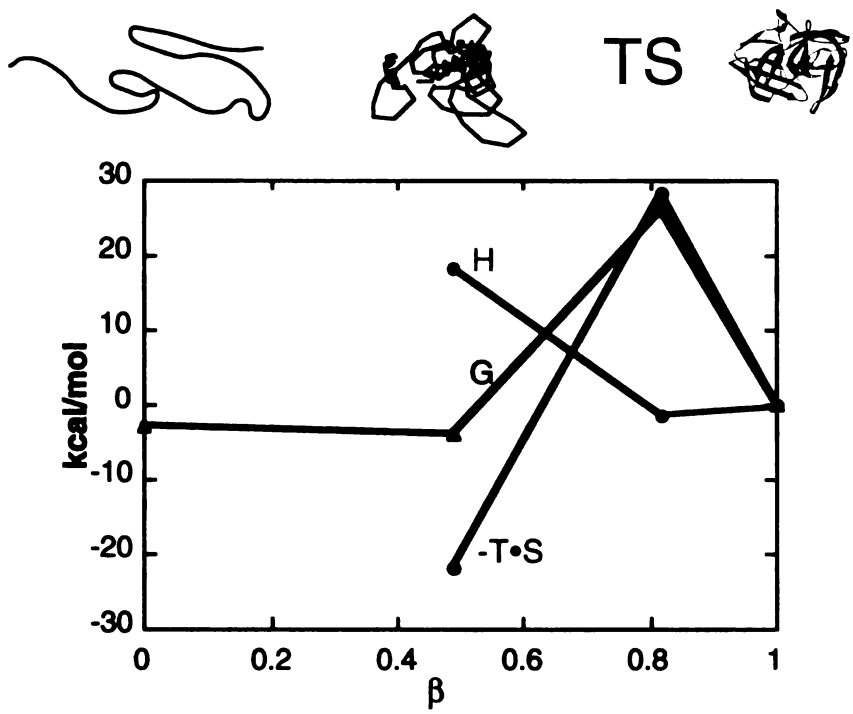


Figure 4.4 αLP folding reaction profile. The change in G, H, and -T•S with position in folding (β), with the native state arbitrarily set to zero. αLP faces an entirely entropic barrier in both folding and unfolding.

In comparison to other proteins, both the enthalpy and entropy of the transition state are unusual. In most other cases we are aware of, at low temperatures there is an enthalpic penalty for both unfolding and folding to the transition state. This is thought to be due to a loss of many favorable native state interactions in unfolding, and the requirement for energy to form new bonds in folding to the transition state (Oliveberg et al., 1995). For αLP at 10 °C, however, the enthalpy of unfolding is slightly favorable (-2 kcal/mol), and there is a -20 kcal/mol favorable enthalpy for folding. While entropic

barriers to folding are common, unfolding is usually accompanied by a loss in entropy (favorable), which is attributed to the gain in configurational entropy.

What is the origin of the favorable enthalpy of unfolding?

One possibility for the absence of an enthalpic penalty for unfolding is that the TS is highly native-like, unfolding only 18% based on the m value analysis, and therefore might retain most of the favorable interactions of the native state. However, CspB, which has an even more native-like TS (<10% unfolded), displays an enthalpic penalty of more than 5.5 kcal/mol between 10 – 37 °C (Schindler and Schmid, 1996). In addition, the slightly favorable activation enthalpy for αLP unfolding is particularly unusual given the substantial increase in heat capacity (1.3 kcal/mol•K) of the transition state. This indicates significant new exposure of hydrophobic surface, which is typically thought to require the breakage of bonds (resulting in a positive, unfavorable enthalpy).

It is possible that the unfolding of αLP occurs through some unusual cooperative transition associated with little enthalpic loss. In addition, the loss of enthalpy from interactions broken in the native state may be compensated by favorable interactions of the solvent. Favorable enthalpy could come from H bonding between newly ordered solvent molecules around the hydrophobic surface, as well as from interactions of the solvent with newly exposed polar groups in the transition state. αLP is highly charged (pI = 10), and there is preliminary evidence that αLP may bury more polar surface area than average (Jaswal & Agard, unpublished data), which could lead to a greater percentage of polar surface available in the transition state to hydrogen bond with the solvent. In any case, it seems that the relationship between the hydrophobic and polar surface exposed in

the transition state remains an important issue to sort out in order to truly understand the origin of the favorable unfolding enthalpy.

What is the origin of the favorable enthalpy of folding?

For most proteins, folding results in an enthalpic penalty over an experimentally accessible temperature range. Some proteins (CI2 and CspB) do demonstrate a large enough decrease in this penalty with temperature, that it leads to an increasingly favorable enthalpy of folding (above 37 °C for CspB (Schindler and Schmid, 1996), and above 50 °C for CI2 (Oliveberg et al., 1995)). Oliveberg and colleagues attribute negative activation enthalpies of folding to "heat capacity changes in the activated complex, temperature-induced changes of the ground state, or a combination of both." (Oliveberg et al., 1995)

It may not be necessary to invoke such changes in the case of αLP, given that there are several possibilities for favorable enthalpic contributions to the transition state. As mentioned before, the αLP transition state is highly native like and therefore may have formed a significant number of favorable internal interactions. In addition, as described in the previous section, enthalpically favorable interactions between ordered solvent molecules, as well as many favorable polar interactions with solvent may be playing a role. The fact that αLP has a negative activation enthalpy of folding at much lower temperatures than other proteins deserves notice, but is not immediately interpretable.

What is the origin of the unfavorable entropy of unfolding?

As has been noted before interpretation of the absolute values of $\Delta S^{\ddagger}$ is tricky, because they "depend on the choice of the preexponential factor in the Eyring equation" (Schindler and Schmid, 1996), which is a matter of continuing debate. However, it's clear that an overall unfavorable entropy results from the dominance of the entropy of solvent ordering around the exposed hydrophobic surface over the favorable entropy of increased configurational freedom on unfolding (Schindler and Schmid, 1996), (Chen and Matthews, 1994). Once again, the role of the solvent in determining the thermodynamics of the transition state is apparent.

Conclusions

Folding of αLP through the intermediate and transition states to the native state involves the burial of 50%, then 80% respectively of the total surface buried in the native state. At 10 °C, folding from the entropically stabilized intermediate to the transition state is accompanied by a huge entropic penalty, likely resulting from the loss of a greater than usual configurational entropy. Such an entropic penalty is typical in the folding of proteins, however, the accompanying gain in enthalpy on folding to the transition state is usually only seen at higher temperatures. This is likely to result in part from the formation of many favorable interactions in the highly native-like transition state.

In addition, we propose that solvent interactions due to the αLP transition state surface exposure may also play a significant role in the favorable enthalpy. The exposure of hydrophobic surface in the transition state, which is ~40% that of the completely unfolded molecule causes ordering of solvent, may lead to favorable hydrogen bonding

between solvent molecules. In addition, the transition state may be stabilized enthalpically by interactions of exposed polar surface with solvent.

Because of these solvent interactions leading to favorable enthalpy, unfolding to the transition state also is favored uncharacteristically by enthalpy. Therefore the barrier to unfolding, as with refolding, is entirely entropic in origin. We ascribe this partly to the solvent effects; as well, the unfavorable entropy of hydrophobic solvation compensates for the favorable gain in conformational entropy.

Additionally, we propose a qualitative model for the transition state in which the two domains of αLP remain largely intact and folded, but open up slightly from each other, to expose significant surface to solvent in the form of a crevice between the two domains (Figure 4.5). This would explain how the large exposure of hydrophobic surface on unfolding to the αLP transition state occurs without a large enthalpic penalty, and with only unfavorable entropy. The two domains remain well packed, retaining the enthalpic stabilization, and since the loosening of the structure occurs without large-scale unfolding, there is not a large enough increase in configurational entropy to outweigh the unfavorable entropy of ordering the solvent along the domain interface.

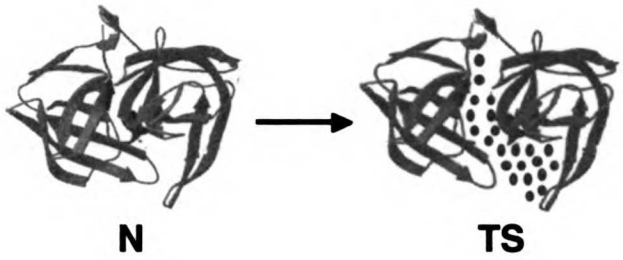
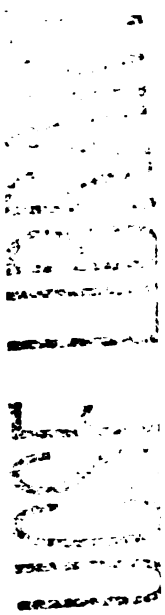


Figure 4.5 Model for αLP TS. Unfolding to the TS occurs through the opening of the two domains, exposing a large amount of hydrophobic surface area to solvent (•) without significant perturbation of the two domains.

Future Directions

Obviously the interpretation of these thermodynamics in terms of a more quantitative physical model for changes in the protein along the reaction coordinate is impossible without knowing the contribution of solvent interactions, both with the protein, and between solvent molecules. As a next step, we plan to use data on the thermodynamics of transfer from oil to water of model compounds approximating the various protein groups to calculate the estimated solvent contributions to the thermodynamics of unfolding (Makhatadze and Privalov, 1995). This will allow us to roughly deconvolute the solvent from the protein contributions, allowing actual insight into the internal thermodynamics of the protein.

In order to determine a more accurate estimate for the activation $\Delta G^{\ddagger}$ and $\Delta S^{\ddagger}$, it would be useful to calculate them based on the other extreme estimate for the preexponential factor, which is on the order of 10^6, based on estimates of diffusion-limited time constants for folding processes ranging from nucleating a helix to forming a 6-10 residue loop (Arrington and Robertson, 1997). By measuring the dependence of



folding and unfolding on viscosity (Jacob et al., 1997), it may be possible to estimate the contribution of diffusion of the polypeptide chain to the entropy of folding and unfolding.

Determination of the activation volume of the transition state through pressure dependence measurements would be useful (Pappenberger et al., 2000). This would help to test the model of a highly compact transition state with a possible deep solvent exposed crevice between the two domains leading to a high $\Delta C_p^{\ddagger}$ without loss of enthalpy upon unfolding to the transition state. If that model is correct, the volume difference between the native and transition state should be very small.

Given the suggestion that polar interactions with solvent may stabilize the transition state enthalpically, exploring the pH and ionic strength dependence of the unfolding barrier would be extremely interesting. Although experimentally challenging, due to the sensitive nature and laborious process for measuring refolding in the absence of pro region, it would be very interesting to determine the temperature dependence of the refolding barrier, even over a limited temperature range. This would provide insight how the enthalpy and entropy change from what we have determined here at 10 °C. In addition, determining the thermodynamics of the refolding barrier under any interesting pH or salt conditions found in the unfolding experiments would help elucidate the electrostatic nature of the barrier.

It is likely to be experimentally much easier to explore refolding conditions with the R102H, G134S mutant, which refolds 400 times faster than the wildtype (Derman and Agard, 2000). Preliminary experiments monitoring the unfolding rate of this mutant suggest that most of the ability to refold faster is due to an effect of stabilizing the transition state (Kelch and Derman, unpublished data). Obviously it would be useful to

follow the mutant's unfolding behavior over a similar range of temperature and denaturant and compare the thermodynamics of the unfolding barrier with those described here, in order to determine how much they are changed.

Although the eventual deconvolution of the solvent and protein effects will give a much more detailed picture of what is going on with the protein than is usually possible, it will not lead to an understanding of the structural details of the transition state. The method of phi value analysis, developed in the Fersht laboratory, has been fairly successful in identifying residues which are natively structured in the transition state, by comparing the effect of the single mutations on the rates of folding or unfolding compared to the effect on the equilibrium unfolding (Matouschek and Fersht, 1991). For αLP this would entail probably using the faster folding mutant, and measuring refolding and unfolding profiles of mutants. Based on the structure of the pro region native state complex and the resulting folding model, possible important regions contributing to the folding barrier, such as the beta hairpin, could be mutated (i.e. alanine scanning) and used in phi value analysis.

The highest resolution understanding of the structural transitions in unfolding to the transition state will result from performing Native State Hydrogen Exchange experiments in low concentrations of denaturant on both αLP. This method allows the detection of very rare partially unfolded intermediates that are sampled at equilibrium (Chamberlain et al., 1996), (Bai, 1995). Regions of the protein which unfold transiently, even with a frequency as low as 1 in 10^6 (Chamberlain et al., 1996), are labeled by the exchange of amide protons for deuterium, which can be detected by the change in the 2D NMR spectrum of the native protein.

In the case of αLP, these experiments will reveal only intermediates sampled between the native and the transition state, since any molecules unfolding past the transition state will not be able to refold to the native state to allow detection (because the unfolded conformations are more stable than the native state (Sohl et al., 1998)). It will be interesting to compare the intermediates and their relative destabilization in the faster folding mutant compared to wildtype αLP.

Chapter 5. Complex Unfolding Kinetics Reveal Aggregation Behavior of the αLP Intermediate.

Introduction

At 4 °C, unfolding of αLP is clearly two state, with virtually identical rates for the loss of tertiary structure, monitored by fluorescence, and loss of secondary structure, monitored by circular dichroism (Figure 5.1). This indicates that the loss of structure during αLP unfolding is highly cooperative and takes place in single unfolding transition. This supports the use of fluorescence, which requires minimal protein (0.1 – 2 μM) and allows four different unfolding reactions to be monitored simultaneously, through the use of a 4-cell automated turret available for the fluorimeter, for the measurement of unfolding kinetics.

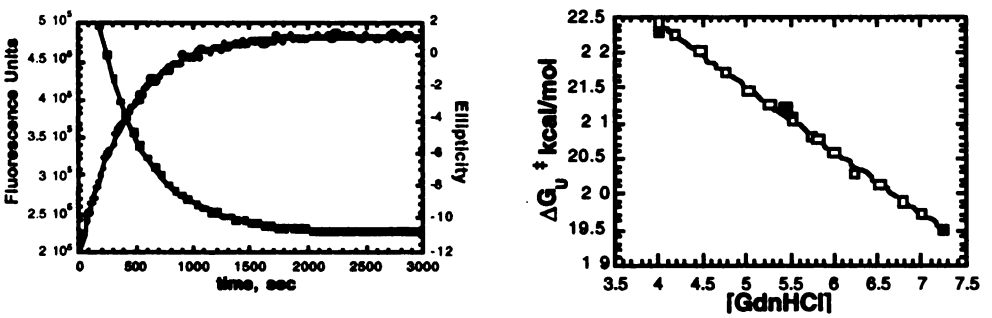
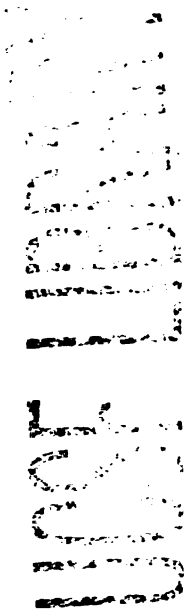


Figure 5.1 SA195 αLP unfolding monitored by CD and fluorescence at 4°C. a)Unfolding in 7.25 M GdnHCl displays a cooperative loss of fluorescence (blue) and ellipticity (red) with a rate of 400 sec^{-1} b) CD and Fluorescence rates have the same GdnHCl dependence.

Experiments monitoring unfolding at higher temperatures in the absence of denaturant revealed that the correspondence between fluorescence and circular dichroism kinetics breaks down under these conditions. Although the unfolding processes



monitored at such high temperatures are not likely to have physiological relevance, measurement of rates up to 70 °C were necessary to carry out the determination of thermodynamic components of the unfolding barrier and the analysis of the denaturant binding behavior. Therefore it was important to confirm that the rates measured by fluorescence under these conditions were still correctly reporting on the unfolding barrier between N and I. Preliminary experiments detailed in this chapter demonstrate that this is the case, and lead to a possible model for the interpretation of the complex unfolding kinetics at high temperatures in the absence of denaturant in terms of properties of the intermediate state and the unfolding landscape of αLP.

Methods

Fluorescence. Using an excitation wavelength of 283 nm and monitoring emission at 322 nm, the decay in fluorescence of samples at either 1.75 μM or 5 μM was measured on an Aminco 8100SLM Spectopolarimeter. For each unfolding reaction, 290 μl 10 mM KOAc pH 5.0 solution (with appropriate GdnHCl concentration) lacking protein was equilibrated in the fluorescence cuvette at the appropriate temperature for at least 10 minutes before unfolding was initiated by addition of 10 μl protein stock followed by mixing.

Circular dichroism. For each unfolding reaction, 393 μl 10 mM KOAc pH 5.0 solution (with appropriate GdnHCl concentration) lacking protein was equilibrated in a 2 mm cuvette at the appropriate temperature for at least 10 minutes before unfolding was initiated by addition of 7 μl protein stock to 5 μM final concentration, followed by mixing. The change in elliptical absorbance in time scans at a single wavelength

(between 205 – 225 nm) or in wavelength scans from 200 – 260 nm were measured a Jasco J715 spectropolarimeter with a Peltier temperature controller.

ANS binding. ANS was made up at 10 mM in EtOH, then diluted to 1.5 mM in 10 mM KOAc pH 5.0, and used at 50 μM in the unfolding reactions. ANS binding was followed by monitoring fluorescence emission at 475 nm after excitation at 380 nm.

Unfolding monitoring inactivation of wildtype protease. Wildtype αLP was added to 10 mM KOAc pH 5.0 preequilibrated in a waterbath at 70 °C, at a final concentration of 4.85 μM. Aliquots were removed over time and assayed for activity on a paranitroanilide substrate.

Data Analysis. Data were fit to a single exponential decay using Kaleidagraph.

Results

Experiments exploring the temperature and denaturant binding dependence of unfolding described in Chapter 3 and 4, led to the discovery that at 60 and 70 °C it is possible to directly observe unfolding in the absence of denaturant by fluorescence. However, separate experiments monitoring unfolding at 80 °C by CD had yielded a rough rate of unfolding slower than that observed by fluorescence at 70 °C. To investigate this apparent discrepancy, I directly compared the rates of structure loss by CD and fluorescence at 60 and 70 °C (Figure 5.2). I observed that the rate by CD was in fact slower by 20-fold at 60 °C and 60-fold at 70 °C.

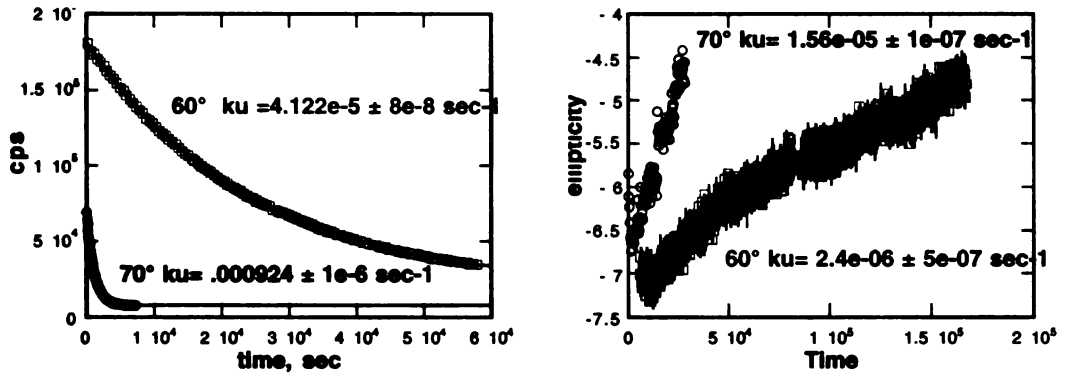


Figure 5.2. Unfolding by CD and fluorescence at 60 and 70 °C.

This confirmed that there is a discrepancy between kinetics monitored by fluorescence compared to CD under these conditions of high temperature and lack of denaturant. One possible explanation is that at 5 μM (used for CD), SA195 is more prone to association than at 1.75 μM (used for fluorescence) and populates an oligomerized state which would unfold more slowly than the monomer. To test whether the difference was simply an effect of the concentration, I measured the loss of fluorescence of a 5 μM sample at 70 °C. While that rate was slightly slower (0.000601 sec^{-1} ± $6x10^{7}$) than the average of two separate measurements of 1.75 μM fluorescence decay at 70 °C (0.00085 ± 0.0001) it does not account for the 60-fold difference in rate observed by CD.

An alternate possibility is that at higher temperatures in the absence of denaturant, unfolding becomes less cooperative, and fluorescence is monitoring a lower energy local unfolding event involving the buried tryptophan, while complete unfolding occurs at the slower rate followed by CD. If this is the case, there should still be secondary structure present in a sample apparently completely unfolded by fluorescence. To test this, I took the 5 μM sample whose fluorescence decay was complete at 70 °C and monitored its CD

signal over time. That signal slowly decreases, as shown in Figure 5.3a, with slightly different kinetics compared to a second 5 μM SA195 sample at 70 °C monitored solely by CD from the initiation of unfolding through addition of the protein to buffer at 70 °C.

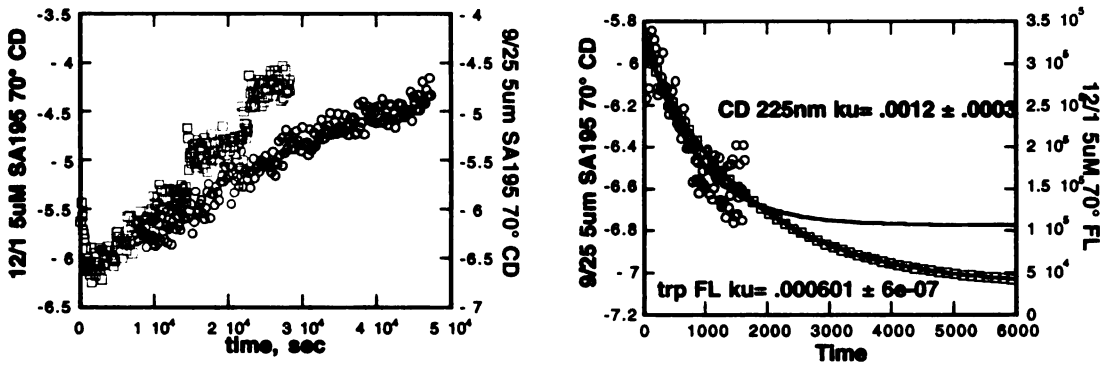
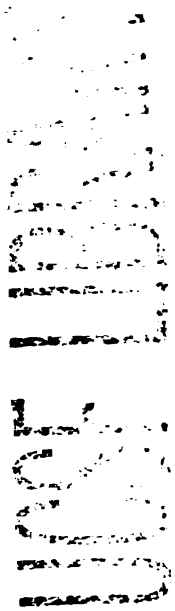


Figure 5.3. 70°C unfolding of SA195. a) CD and b) initial CD compared to fluorescence.

Closer examination of the beginning of the CD time trace at 70 °C of that second sample reveals an initial increase in signal of about 1 mdeg over the first 2000 seconds, before the signal begins to decrease. Fitting just the initial 2000 seconds of the CD signal at 70 °C gives a single exponential rate constant of 0.0012 ± 0.0003 sec^{-1}, which is within two-fold of the rate of fluorescence loss (Figure 5.3b). This result suggested that the same unfolding event involving a loss of fluorescence is detected by CD, but as an unexpected increase in ellipticity. The slow loss in ellipticity is due to a second transition, not involving a change in tertiary structure.

But what unfolding transition leads to a loss of tertiary structure by fluorescence while increasing ellipticity at 222 nm by CD? A clue to the identity of the mysterious state resulting from that transition came from looking back at the normalized comparison of native, intermediate and unfolded CD signals (Baker et al., 1992b). At 222 – 225 nm, the intermediate state of αLP actually has more ellipticity than the native state. In fact,



not only does the intermediate display more signal at 4 °C, but its ellipticity at 222 nm increases with temperature (Figure 5.4), so that at 70 °C, it's reasonable to propose that the initial increase of ~1 mdeg is due to unfolding from the native to the intermediate state.

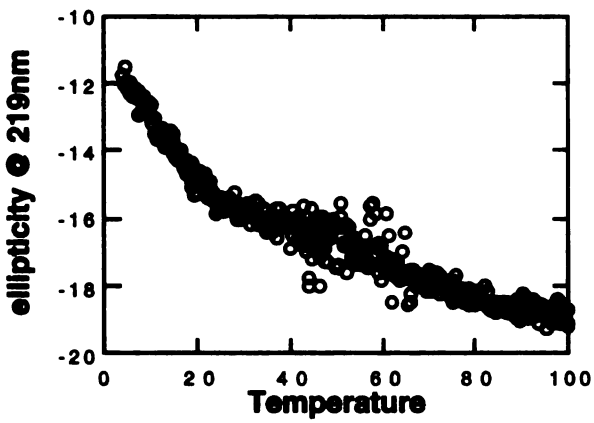
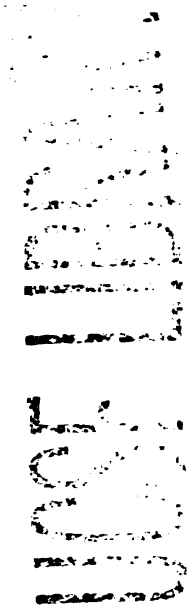


Figure 5.4. Change in intermediate ellipticity with temperature.

These results offered a possible explanation for the complex unfolding kinetics under these conditions. What fluorescence and the initial increase in CD signal were monitoring was the unfolding of the native state to intermediate (1), while the subsequent decrease in CD signal was monitoring some rearrangement or aggregation process of the intermediate state (2).



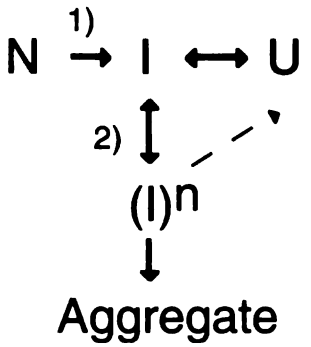


Figure 5.5 Model for unfolding at high temperatures in the absence of denaturant.

This possibility was supported by successive wavelength spectra of a native SA195 sample at 80 °C in the CD (Figure 5.6).

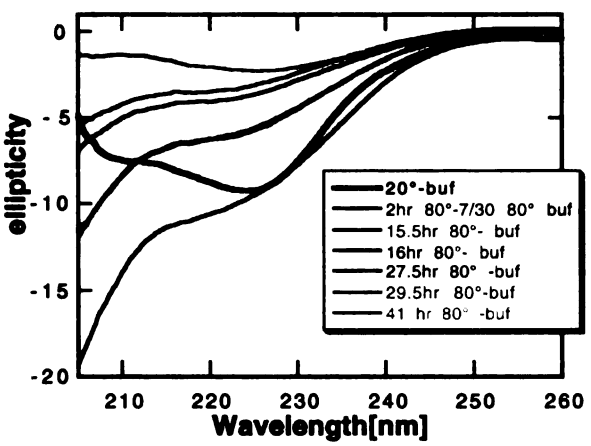


Figure 5.6 Change in CD spectrum of native SA195 during unfolding at 80 °C.

After two hours, the spectrum in red is still very similar to the native spectrum in black above 225 nm, but below that the ellipticity has increased a huge amount. Later spectra at 16, 28, and 41 hours show the slow decrease in the entire signal. Since the difference between the native and intermediate CD spectra got much bigger below 222 nm, where I had previously monitored the initial increase of only ~1 mdeg in ellipticity, I

monitored another 5 μM sample at 70 °C by CD but at 215 nm. Indeed, the increase was much larger, at 4.5 mdeg, and fit to a rate of 0.00197 ± $3x10^5$ (Figure 5.7).

As an additional probe of native structure, I followed inactivation of the wildtype enzyme through autoproteolysis, which matches the rate of global unfolding of the inactive enzyme within two-fold (see Chapter 6). Loss of activity at 70 °C occurred with a rate of 0.00154 ± $3x10^5$ sec^{-1} (Figure 5.7), very close to the initial CD increase at 215 nm. This provides further evidence that the first fast rate being observed by fluorescence and CD corresponds to the loss of native structure.

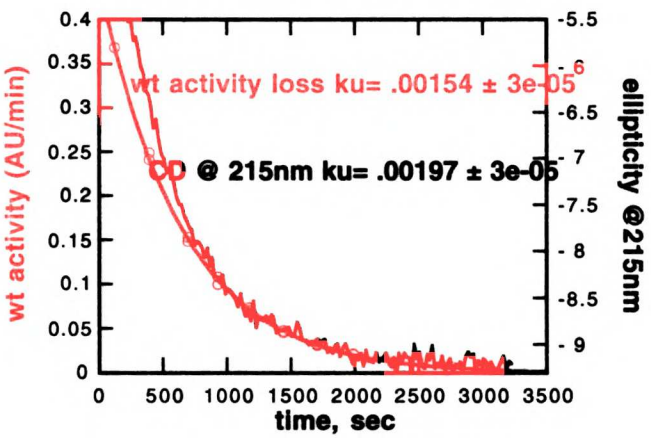


Figure 5.7. Unfolding of αLP at 70 °C by initial CD ellipticity increase and loss of wildtype activity.

Although the increase in CD could be interpreted as due to the appearance of the intermediate, I wanted to use an additional method known to confirm the appearance of the intermediate. As described in chapter 2, the αLP intermediate binds ANS, which is a hallmark of molten globule intermediates, since it only binds states exposing hydrophobic surfaces, not native or unfolded states. The results of one experiment in which I

alternately monitored the loss of intrinsic tryptophan fluorescence and the increase in ANS fluorescence during unfolding of a single 1.75 μM protein sample at 70 °C in the fluorimeter are shown in Figure 5.8.

ANS fluorescence does increase, indicating appearance of the intermediate state, however the rate is about six-fold faster than the rate of loss of native fluorescence. It is hard to imagine a possibility whereby the intermediate forms, giving rise to ANS binding, but somehow has a conformation that retains tertiary structure displaying a native signal. Given the minimal data recorded following the ANS fluorescence, it is not clear that there is a real difference. However, it could reflect in part the fact that the αLP native state does actually display some ANS fluorescence (Figure 2.1).

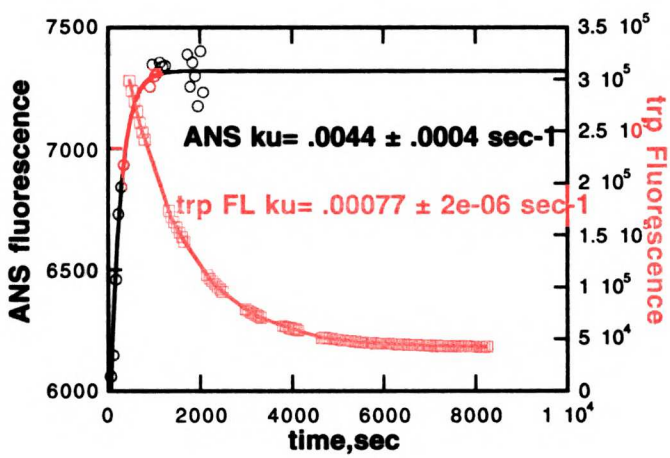


Figure 5.8. Unfolding of αLP at 70 °C followed simultaneously by trp and ANS fluorescence.

The results of all the experiments described above measuring unfolding at 70°C in the absence of denaturant are summarized in Table 5.1. There are four rough groupings of rates determined by the various techniques. The fastest rate was measured for ANS

binding, at 0.004 sec^{-1}. Monitoring the increase in CD ellipticity at either 222 or 215 nm, and the inactivation of the wildtype protease gave rates two to three-fold slower than ANS binding, at 0.0012 to 0.002 sec^{-1}. Tryptophan fluorescence reported rates slower by one and a half to three-fold slower, at 0.00085 sec^{-1} for 1.75 μM and 0.0006 sec^{-1} for 5 μM protein. Slowest by up to two orders of magnitude were the rates of ellipticity decrease.

Table 5.1 Rates of conformational change at 70°C.

Technique	[protein], μM	ku (sec^{-1})
ANS fluorescence	1.75	0.0044 ± 0.0004
CD increase, 215 nm	5	0.00192 ± 0.00003
wildtype inactivation	5	0.00154 ± 0.00003
CD increase, 222 nm	5	0.0012 ± 0.0003
Trp fluorescence	5	$0.0006012 \pm 6x10^{-7}$
Trp fluorescence	1.75 (ave of 2)	0.00085 ± 0.00001
CD decrease, 222 nm	5	$1.56x10^{-5} \pm 1x10^{-7}$
CD decrease, 222 nm post fluorescence	5	$2.67x10^{-5} \pm 5x10^{-7}$

Discussion

The following model for unfolding at 70 °C in the absence of denaturant based on the separation of the techniques into those reporting on the N to I transition, and the slower CD signal decrease.

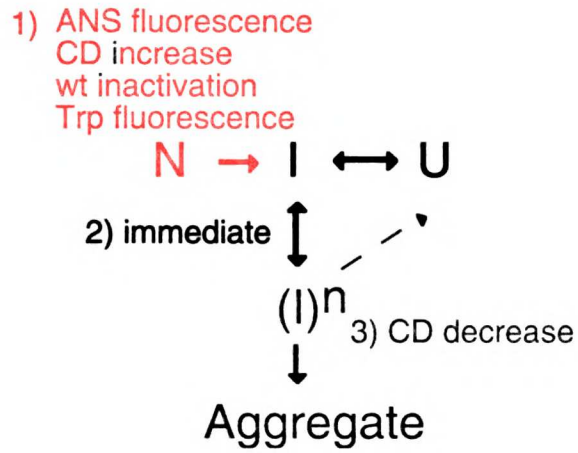


Figure 5.9 Model of unfolding and rearrangement at 70 °C. 1)N unfolds to I with a rate between 0.004 – 0.0006 sec^{-1} is detected by both techniques following the loss of N and the appearance of I, 2) I associates immediately to form I^n, 3) I^n aggregates or unfolds with a rate of 2 – 3 $x10^{-5}$ sec^{-1}.

In these rates being taken as reporters of the N to I transition there is some spread: at most 7-fold between ANS binding and 5 μM tryptophan fluorescence. However, the ANS binding needs to be repeated with more careful monitoring, and with concentration matched controls of the intermediate and native ANS binding levels before that rate can be strictly believed. Without that rate, there is at most a 3-fold spread in the values, corresponding to free difference of only 0.8 kcal/mol.

One possibility for variation between experiments leading to such a spread is the temperature. The stability of the water baths used to maintain 70 °C during the fluorescence and inactivation experiments is supposed to be ± 0.2 °C. However, there is tubing between the bath supplying the fluorimeter with the heated liquid, and that is supplied to the jackets around the cuvettes. In monitoring the actual temperature of the cuvettes compared to the temperature displayed by the bath, there can be up to 0.5 °C variation. The Peltier temperature control unit on the CD is stable to ± 0.1 °C, but again there is a significant length of tubing connecting it to the sample chamber. In addition, for technical reasons when the protein is being added to the preequilibrated buffer in the cuvettes, the entire cuvette and the sample chambers are exposed to room temperature for

15 - 30 seconds. At this high temperature even a 0.5 °C difference can have large effects on the rate.)

The most important issue to resolve is whether the unfolding barrier under these conditions of high temperature and no denaturant is losing its cooperative nature. Our initial concern, that the slow decrease in CD indicated the true rate of global unfolding of N, can be ruled out. Regardless of the spread, the combination of faster rates measured by tryptophan fluorescence, ANS binding, and initial CD increase provide convincing evidence that native tertiary structure and activity are being lost with a rate roughly corresponding to the rate of appearance of the intermediate. Therefore using tryptophan fluorescence to monitor unfolding over a range of temperature and denaturant conditions is warranted.

The complex kinetics that are observed by CD are independent of the unfolding barrier, and result from properties of the intermediate that become revealed under these conditions. It is likely that in the absence of denaturant, which completely unfolds the intermediate at concentrations as low as ~0.75 M, the intermediate is stabilized over the unfolded state, and can be observed.

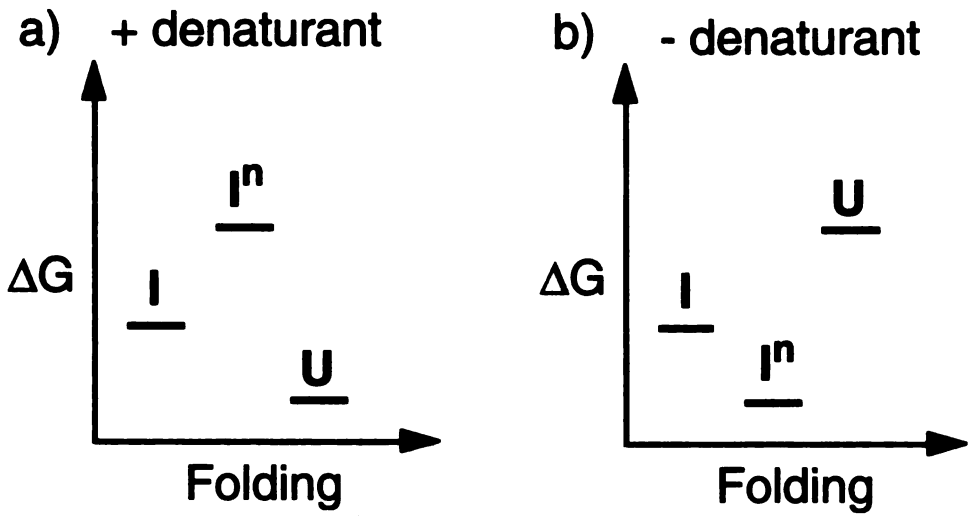


Figure 5.10 Effect of denaturant on relative stability of I, I^n, and U. a) In the presence of denaturant, U is more stable than I so unfolding occurs from N to U. b) In the absence of denaturant, I is more stable than U, and I^n is the most stable, so unfolding occurs from N to I which associates to form I^n.

From attempts to thermally melt the intermediate, we knew that it displays unusual behavior at high temperatures including a tendency to aggregate and increased ellipticity. Therefore because of the high temperature, as soon as the native state unfolds to the intermediate, it undergoes a rapid association to some oligomerized or even aggregated state, which results in the increased ellipticity. Increased β sheet content as detected by CD is known to accompany aggregation of proteins (Fink, 1998).

It would be interesting to monitor unfolding by CD at 60 °C in the presence of denaturant, to see where the correspondence between fluorescence and CD breaks down. The prediction would be that above the denaturant concentration needed to unfold the intermediate, a single exponential decay in ellipticity matching fluorescence as the native state unfolds directly to the unfolded state would be observed. Below that concentration,

either I or I^n should be detectable, depending on the effect of the low concentration of denaturant on the association to form I^n.

It is unclear which process is then giving rise to the slow loss of ellipticity. It could be due to the slow unfolding of this associated intermediate to the unfolded state. However a more likely possibility is that it is simply due to the loss of soluble protein through precipitation from this associated state. Some protein precipitate was observed in the 70 and 80 °C cuvettes. The spectrum of the sample at 80 °C after 41 hours does appear to have some shape rather than the characteristic random coil pattern, which might indicate that it results from a very low amount of protein still soluble. This could be tested by spinning down a post 70 or 80 °C CD sample that has lost ellipticity, and comparing the A_{280} of the supernatant to the beginning A_{280}. In addition, oligomers could be detected by doing mass spec on the post 70 or 80 °C CD samples. Aggregation could be directly followed by monitoring the OD_{340} of the protein at 70 °C.

With the probable demonstration that these high temperature samples are undergoing aggregation, it would be of interest to determine whether this was random aggregation or ordered association. Given the known propensity for β-sheet proteins to associate and the demonstration that several non-disease proteins can form amyloid fibrils (Dobson, 1999), it is very likely that these conditions have begun to reveal that corner of αLP's folding landscape. Post 70 °C protein samples could be looked at for possible fibrils using EM. Congo Red and Thioflavin T binding (probes of fibril formation) could be monitored at 70° C to observe kinetics of fibril formation.

Amyloid formation by αLP would be particularly interesting as another example of the importance of kinetic stability its landscape. While the kinetic barrier separating N

and I has been well determined, the possibility for fibril formation would point to the existence of another very high and very important barrier, between I and I^n. Determining the energetic and structural features of the barrier to fibril formation would provide insight into the provocative question of how nature has managed to evolve two separate high barriers preserving the biological activity and structural integrity of the enzyme.

Chapter 6. Evolution of the perfect protease.

Introduction

Proteases play ubiquitous and critical roles in biological systems, ranging from digestion of proteins to provide nutrients to regulating signal transduction and cell cycle control (Neurath and Walsh, 1976), (Neurath, 1984). While their catalytic mechanisms have been well studied, little is know about how highly active, constitutive proteases have themselves evolved to avoid proteolytic destruction. Experiments described here on the extracellular bacterial protease α-lytic illustrate how optimal resistance to proteolysis is achieved through extreme restriction of global unfolding transitions and suppression of local dynamics.

Intuitively, minimizing susceptibility to proteolysis should involve reducing the number of flexible and disordered loops on the target protein as well as optimizing the choice of surface exposed residues. Mechanistically, this derives from the fact that proteases typically require between four and eight residues of the substrate peptide chain to fit in a precise manner into the deep cleft that forms the active site (Perona and Craik, 1995). Thus only flexible regions of a protein target would be able to adopt conformations required for proteolysis. However, it must be recognized that even very resistant native states are conformationally dynamic and constantly sample open conformations, in processes ranging from small breathing motions to global unfolding. Assuming a protease resistant native state, proteolytic degradation of a protein can be described as follows (Rupley, 1967):

Figure 6.1 Conformational opening reaction leading to proteolytic cleavage.

In order to avoid proteolytic attack and prolong its lifetime, a protease should restrict those unfolding and dynamic opening events leading to protease accessibility. Obviously some proteases are proteolytically cleaved as part of their physiological maturation and regulation, so their native states might purposefully contain a proteolytically sensitive loop. Many proteases are meant to be turned over at some physiologically necessary rate through proteolytic degradation, and therefore the rates of their sampling of open conformations have been evolved to allow that.

For proteases in certain environments, however, like those which are extracellular or in contained compartments of the cell dedicated to degradation, there is no need for regulated turnover, and they may provide the best examples of how proteolysis can be resisted. A common feature of a broad class of such proteases, encompassing virtually all extracellular bacterial proteases across different evolutionary families, as well as lysozomal and vacuolar eukaryotic proteases, is their synthesis with a pro region. Pro regions are segments of polypetide chain, ranging from 40 to ~550 residues either N- or C- terminal to the mature protease (Baker et al., 1993). Where they have been investigated, such pro regions appear to play a central role in facilitating the folding of

their attached proteolytic domains before being proteolytically processed and subsequently degraded.

One of the best studied examples of pro region dependent folding is the extracellular α-lytic protease (αLP) of *Lysobacter enzymogenes*. The protein folding energetics of αLP are vastly different from those of most proteins. The native folded αLP is less thermodynamically stable than both the unfolded protein and a molten globule intermediate in its folding pathway (Sohl et al., 1998). In addition, while a typical small protein folds in μs to seconds (Jackson, 1998), folding of the αLP occurs only with a half time of ~3700 years. In the presence of the pro region, however, folding occurs in seconds, because binding of the pro region both catalyzes folding and stabilizes the native state. Once the protease is folded, the pro region is degraded, leaving the protease trapped in the metastable native conformation. Therefore in contrast to most proteins, it is not thermodynamic stability, but kinetic stability, determined by an enormous barrier to unfolding ($t_{1/2}$ ~ 9 months at 25 °C), that preserves αLP in its enzymatically active native state. We investigated the possibility that the high cooperativity of kinetically-determined protein stability may offer a functional advantage over conventional thermodynamic stability mechanisms by severely restricting protein dynamics, leading to enhanced resistance to proteolysis.

Methods

Protein expression and purification

αLP and SA195 αLP were expressed in *E. Coli* and purified as described (Mace and Agard, 1995) and (Sohl et al., 1998) respectively. Rat trypsinogen was expressed in

Pichia as described (Halfon and Craik, 1996). Secreted protein autoactivated to trypsin, which was purified as follows. All buffers contained 10 mM calcium chloride. The supernatant was slowly brought to 4.5 M NaCl, loaded onto a phenyl sepharose fast flow column, washed with 50 mM MES pH 6.0, 4.5 M NaCl then eluted with 50 mM MES pH 6.0. After dialyzing to remove NaCl, the eluted protein was loaded onto a p amino benzamidine agarose column, washed with 50 mM Tris pH 7.5 and eluted with 100 mM acetic acid into fractions containing 1/20 volume 0.5 M sodium citrate pH 4.5. Fractions with active protein were pooled and dialyzed into 1 mM HCl and assayed for activity as described (Halfon and Craik, 1996).

Survival Assays

6.5 μM αLP, trypsin, and chymotrypsin (TLCK-treated, Worthington)were incubated together at 25 °C in 10 mM $CaCl_2$ and 50 mM KOAc pH 5.0, 50 mM MOPS pH 7, 50 mM HEPES pH 7 or 50 mM Tris pH 8. Aliquots were removed over time and the survival of the individual proteases was measured based on their activities, which could be distinguished given their non-overlapping specificities for different substrates (succinyl-Ala-Pro-Ala-pNA, succinyl-Ala-Ala-Pro-Arg-pNA, succinyl-Ala-Ala-Pro-Leu-pNA, used for αLP, trypsin and chymotrypsin, respectively, all at 1 mM in 10 mM $CaCl_2$, 100 mM Tris, pH 8). Cleavage of the proteases during survival assays was monitored by quantitating the disappearance of the full length species as monitored by band intensity in a Coomassie stained SDS PAGE gel, using ImageQuant for Macintosh v1.2.

The inactivation in three separate experiments at each pH, and cleavage in single experiments at pH 5 and pH 7 were fit for each protease with a single exponential equation y = m1 - m2*(1-exp(-m3*m0)). To normalize for slight concentration

variations in separate survival assays, the raw data were treated by subtraction the offset (m1 – m2) and dividing by the amplitude of the decrease (m2). The normalized data from three separate survival assays for each protease at each pH was then combined and fit with the single exponential equation to determine the inactivation rate and standard error given in Table 6.1.

Autolysis Assay

3.25 μM αLP was incubated in 1 M GdnHCl 10 mM KOAc pH 5.0 at 25 °C. αLP activity was followed by monitoring hydrolysis of 1 mM succinyl-Ala-Pro-Ala-pNA in 100 mM Tris pH 8.0 by aliquots over time.

Global Unfolding

Fluorescence measurements of the inactive mutant SA195 αLP at 1.75 μM in 0.98 M GdnHCl were made with excitation at 283 nm, emission 322 nm in an 8100SLM-Aminco fluorimeter, with slit widths set to 4, 0.25, 0.25, 16, 16, connected to an external waterbath set to 25 °C. The sealed fluorescence cuvette was maintained at 25 °C between measurements in a 25 °C incubator.

Results and Discussion

In order to test the hypothesis that kinetic stability offers a functional advantage over thermodynamic stability, we have directly compared the lifetimes of αLP and its thermodynamically stabilized homologues, chymotrypsin and trypsin, under highly proteolytic conditions. In survival assays mixing equimolar amounts of the three proteases at pH 5, 7, and 8, αLP's biological activity remains virtually unchanged, while its mammalian counterparts are readily destroyed at rates ≥ 100-fold faster than αLP

(Fig.6.2, Table 6.1). This is not due to a greater number of or more accessible proteolytic sites on chymotrypsin and trypsin. In fact, normalized to the length of the protein, αLP has the same percent of sites as chymotrypsin with greater than 20% surface accessibility and ~30% more than trypsin. Nor is αLP's longer lifetime due to an ability to remain active when cleaved. As shown in Figure 6.2a for pH 7, the rates of cleavage monitored by gel for all three proteases are within error of their inactivation rates. Therefore, αLP clearly has a superior ability to withstand proteolysis and autoproteolysis for months longer than its thermodynamically stabilized counterparts.

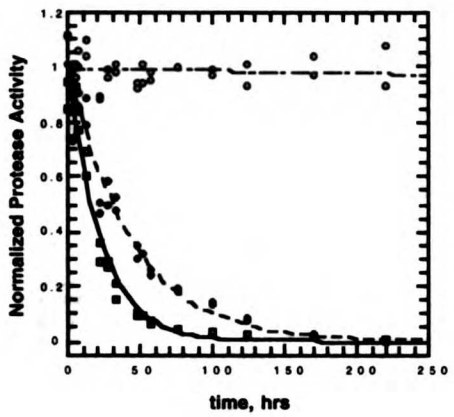
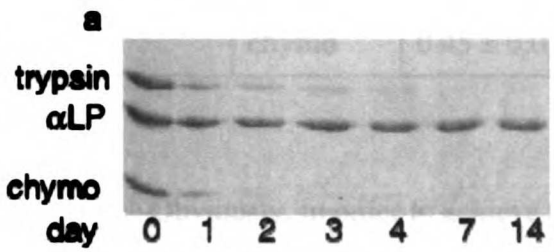


Figure 6.2 Survival Assay at pH 7, 25 °C. αLP (red) is more resistant to proteolysis than chymotrypsin (black) and trypsin (blue) as monitored by loss of proteolytic activity.



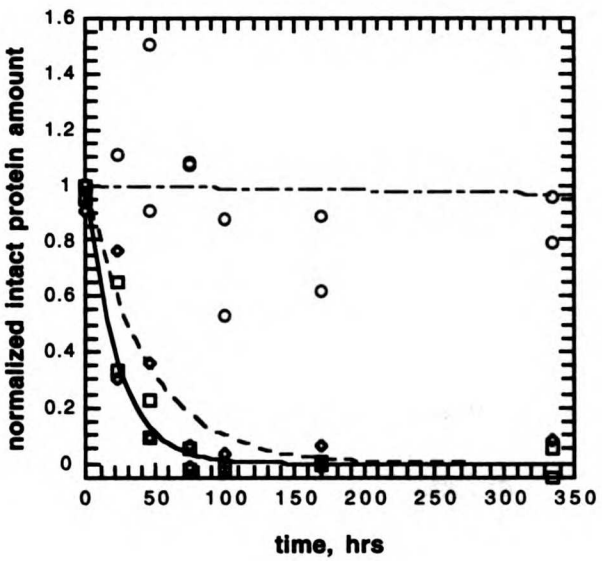


Figure 6.3 Rate of cleavage vs. rate of inactivation. Quantitation of the loss of intact protein monitored by a) SDS PAGE, b) compares closely with the fit curves for the loss of activity.

Table 6.1. Survival assay rates. Rates in $days^{-1}$ by followed by inactivation (loss of activity on substrate) or cleavage (loss of band on gel).

	protease	inactivation	cleavage
pH 5	αLP	220 ± 70	100 ± 25
	trypsin	1.93 ± 0.09	3.5 ± 0.6
	chymo	**17 ± 3**	**14 ± 9**
pH 7	αLP	180 ± 40	110 ± 90
	trypsin	1.21 ± 0.07	1.3 ± 0.3
	chymo	**0.70 ± 0.04**	**1.1 ± 0.2**
pH 8	αLP	300 ± 200	n.d.
	trypsin	1.6 ± 0.1	
	chymo	**0.45 ± 0.01**	

Obviously, in order to achieve this extreme longevity, αLP must be highly successful at preventing the sampling of proteolytically vulnerable locally and globally unfolded states (see equation above). In fact, in the most optimal case, local unfolding in αLP would be suppressed, so that it is exposed to proteolysis only once it has completely unfolded, which has already been shown to occur on a dramatically slow time-scale ($t_{1/2}$ ~ 9 months @ 25 °C) (Sohl et al., 1998). To test this possibility, we compared αLP 's rate of inactivation through autoproteolysis to its rate of global unfolding in the absence of proteolysis (measured using the inactive SA195 αLP) at low denaturant concentration, where local unfolding is promoted (Figure 6.4).

The rate of αLP autoproteolysis is 0.02 day^{-1}, compared to the global unfolding rate of 0.04 day^{-1}. This two-fold difference could very likely be due to a greater stability of the SA195 enzyme due to the substitution of the serine for alanine, but in any case

corresponds to a free energy difference of only 0.3 kcal/mol. Therefore, αLP appears to have optimized its protease resistance to the greatest extent possible, by restricting smaller unfolding events so that they do not significantly lead to proteolytic accessibility.

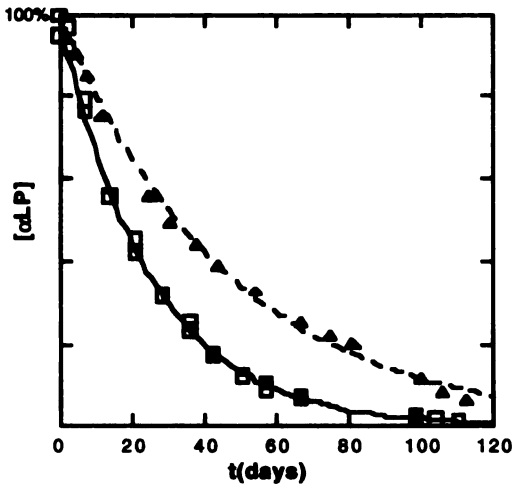


Figure 6.4 Autolysis rate vs. global unfolding rate. Rate of autoproteolysis of the active enzyme (squares) is within two-fold of the rate of global unfolding of the inactive enzyme (triangles).

Using an even more sensitive measure of dynamics, Jonathan Davis and Julie Sohl determined that this restriction of local unfolding is due to an extreme degree of rigidity throughout the entire molecule, not only limiting structural fluctuations leading to protease accessibility but also on the very smallest level of hydrogen exchange. They followed hydrogen-deuterium exchange (HDX) monitored by two dimensional nuclear resonance spectroscopy (2D NMR), and found that αLP displays extremely high protection from hydrogen exchange. While most proteins typically display a "slow exchange core" (Li and Woodward, 1999) of 3 – 8 residues with protection factors

ranging from $10^4 - 10^7$, more than half of αLP's amide protons have protection factors greater than 10^4. Remarkably, one third of those display protection factors greater than 10^9. Nineteen are so protected that no exchange was measured at all even after six months at pH 9, which means that their minimum protection factors are $10^{10.5}$. Not only is this degree of protection among the very highest ever measured in a protein, but the most slowly exchanging residues are not localized to one discrete "core", but are spread throughout both domains of the protein.

This extreme protection, far beyond that seen for typical proteins, suggests that αLP's ability to suppress local unfolding is likely coupled to its high cooperativity of unfolding, since the entire molecule is clamped down so tightly that "local unfolding" does not exist. Because of the cooperativity, a huge amount of energy is required to loosen the native structure at all, and therefore only global unfolding can lead to any accessibility to proteolysis.

As a final indication of the αLP's exquisite optimization, not only has it evolved such cooperativity that the barrier to global unfolding is effectively the barrier to proteolysis, but that global unfolding barrier is at its maximum near the temperature corresponding to the optimal growth range for its bacterial host (27 – 30 °C) (Figure 6.5).

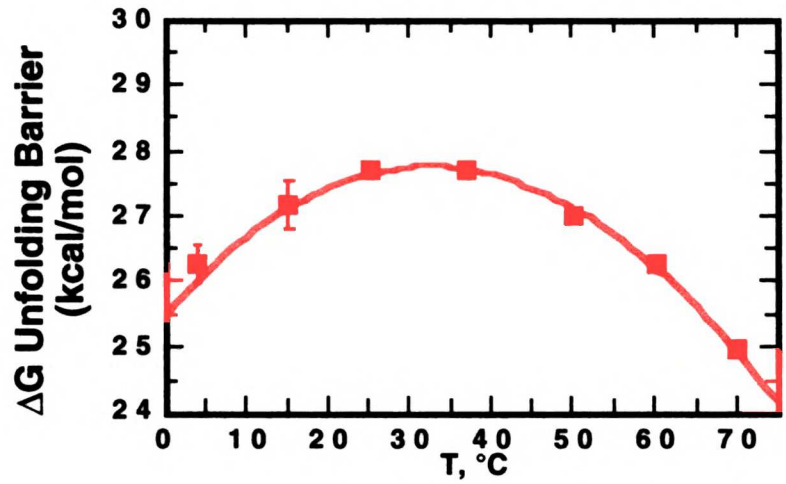


Figure 6.5. Variation in unfolding barrier with temperature.

This high degree of optimization necessary for αLP's extreme longevity has been possible only because the transient existence of the pro region has allowed the independent evolution of the folding and unfolding landscapes. Once folding has occurred and the pro region is cleaved and degraded, the landscape is radically different, which has allowed the folded state to evolve these unique properties to avoid self destruction, despite the consequence of native state metastability. Other pro- containing proteases are likely to have evolved varying degrees of kinetic stability to increase their proteolytic resistance as well. In addition, this strategy of transforming a protein's landscape through covalent modification may be a general one that has evolved to optimize unique functionality in other systems.

Chapter 7. Conclusions: Kinetic Stability in αLP

The experiments on the thermodynamics and kinetics of αLP folding described in this thesis offer a new paradigm for generating extreme stability: kinetic stability facilitated by co-evolution of a pro region to facilitate folding. The defining feature of kinetic stability is the presence of a large barrier to unfolding, preventing the native state from being in equilibrium with unfolded conformations. This has been accomplished through the incorporation of an excess number of glycines, facilitating tight turns and packing, and other mutations leading to a rigid and highly cooperative native state.

What is the role of the pro region in folding αLP*?*

In addition to the glycines causing lower entropy of the native state, they increase the entropy of the intermediate and unfolded conformations of αLP. As a consequence, an excess entropy difference between the native and unfolded conformations leads to the thermodynamic stabilization of the unfolded conformations over the native state. Thus, the barrier to folding is even larger than the barrier to unfolding. Therefore, the pro region performs two roles necessary for guiding folding: 1. to lower the barrier to folding and 2. to stabilize the native state. Because the pro region is highly unstable, once folding of the protease is complete, the pro region is rapidly degraded, leaving the molecule trapped in the metastable native conformation. For the lifetime of the protease, kinetic stability rather than thermodynamic stability maintains it in its biologically active conformation.

What is the origin of kinetic stability?

The intermediate of αLP, which is at the global free energy minimum of the landscape has the characteristics of a typical molten globule ensemble. It is a greatly expanded monomer, exposes hydrophobic patches, and has secondary structure in the absence of tertiary structure. It buries about half of the total surface area buried in the native state. The excess stabilization of the intermediate over the native state is due entirely to entropy, as the enthalpy difference favors the native state. The barrier to folding is also entirely entropic, due to the loss of the high configurational entropy in the intermediate.

The thermodynamic analysis of the unfolding kinetics in the absence of the pro region provides clues as to the nature of the unusually high transition state. It is highly native-like, burying more than 80% of the total surface buried in the native state. Our current model of the transition state consists of the two domains primarily natively folded, but separated enough to create a crevice exposing significant hydrophobic surface (40% of total) to solvent.

Because the two domains independently contain the majority of the stabilizing interactions found in the native state, the transition state has nearly equivalent favorable enthalpy with the native state at low temperatures. In addition, the ordering of solvent around the surface exposed in the crevice could lead to favorable enthalpy from solvent-solvent interactions as well as favorable interactions between solvent and exposed polar surface. The probable exposure of polar surface in the transition state is supported by experiments exploring the temperature and denaturant dependence of the unfolding

kinetics. These reveal that the transition state interacts with GdnHCl through both hydrophobic and polar surfaces, in temperature dependent interactions.

The lack of enthalpy difference between the native and transition states means that the unfolding barrier, as is the case for the folding barrier, is entirely entropic at low temperatures. Because unfolding to the transition state in our model comes from the opening up of the two domains relative to one another, while they retain their intradomain interactions, the increase in favorable configurational entropy should be minimal. Therefore the unfavorable entropy of ordering solvent around the surface exposed in the crevice between domains dominates.

What is the role of pro regions and kinetic stability?

Glycine content appears to be a common feature distinguishing homologous proteases to αLP that have pro regions from those that do not (Table 7.1). The S. griseus proteases, along with several other pro region containing homologues, have 16 – 18% glycines, while the mammalian digestive enzymes and other members of the trypsin serine protease family without pro regions have 7 – 11% glycines. In addition, the β-hairpin implicated structurally as part of the folding barrier is also conserved among the homologues with pro regions.

These common properties may reflect their common function as proteases that break down microorganisms in the extracellular environment, supplying nutrients for their bacterial hosts. The utility of these proteases is compromised by their tendency to degrade themselves as well as other proteins. As such, it is presumably desirable to the

host to evolve proteases that can survive as long as possible under these harsh, degradatory conditions.

Table 7.1 Glycine content of αLP relatives.

Protease	Glycines, no.	Residues, no.	Glycines, %
α-lytic protease family			
L. enzymogenes αLP	32	198	16.2
R. faecitabitus protease I	29	174	16.7
S. albogriseolus protease 20	32	172	18.6
S. fradiae protease 1	31	186	16.7
S. griseus protease A	32	182	17.6
S. griseus protease B	32	182	17.3
S. griseus protease C	35	190	17.9
S. griseus protease D	32	188	17.0
S. griseus protease E	32	183	17.5
S. lividans protease	35	171	20.5
S. livdans protease 0	27	150	18.0
S. spp. alkaline protease I	34	186	18.3
T. fusca serine protease	35	186	18.8
trypsin family			
trypsin	24	224	10.7
chymotrypsin B	23	245	9.4
elastase	27	270	10.0
acrosin	27	436	6.2
achelase 1 protease	25	213	11.7
α-tryptase	19	245	7.8
batroxobin	29	231	8.7
carboxypeptidase A complex III	25	240	10.4
coagulation factor VII	37	406	9.1
collagenase	22	230	9.6
complement factor b	61	739	8.3
enteropeptidase	77	1035	7.4
glandular kallikrein 1	19	238	8.0
granzyme a	22	234	9.4
mast cell protease 7	22	244	9.0
natural killer cell protease 1	19	228	8.3
plasminogen	59	790	7.5
prostasin precursor	28	311	9.0
serine protease hepsin	44	417	10.6

A typical protein stabilized thermodynamically without a large barrier preventing unfolding would constantly sample partially and fully-unfolded states, leading to rapid destruction by exogenous proteases (Fig.7.1a). By contrast, kinetic stability provides a mechanism to increase the cooperativity and raise the barrier to unfolding; thereby suppressing breathing motions and global unfolding (Fig.7.1b). The result is a drastic reduction in susceptibility to proteolytic degradation.

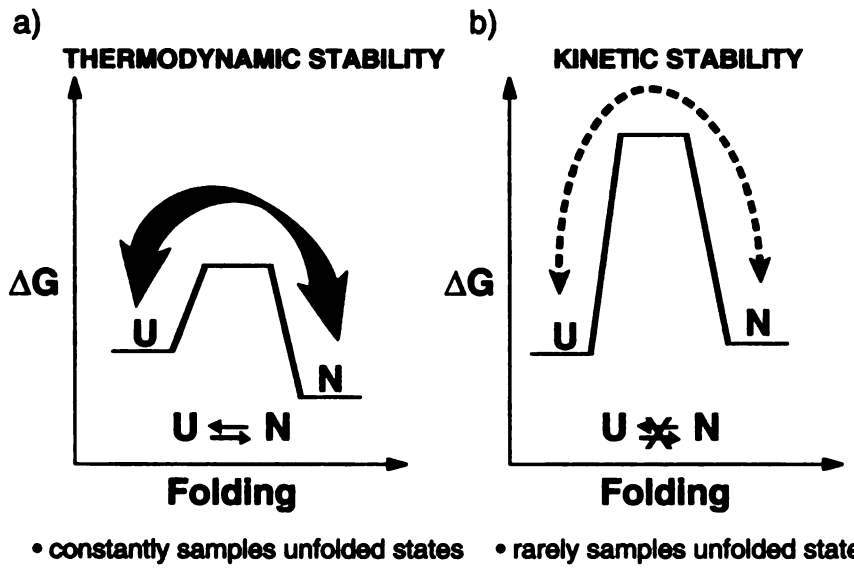


Figure 7.1 Advantages of kinetic stability. a) A typical thermodynamically stable protein without a large barrier rapidly samples fully and partially unfolded states, making it susceptible to proteolysis. b) A kinetically stable protein only rarely samples these unfolded states, making it much more resistant to proteolysis. In the case of αLP, the native state is less stable than the unfolded states, however kinetic stability does not require a metastable native state.

As demonstrated in Chapter 6, this has indeed been a successful strategy for extending αLP's lifetime when compared to its thermodynamically stabilized homologues chymotrypsin and trypsin. In survival assays where these three proteases are mixed and allowed to attack each other, αLP retains its biological activity for much longer than its mammalian counterparts. The sensitivity of trypsin and chymotrypsin to proteolysis is

likely to be a necessary aspect of their regulation *in vivo*. Additionally, the rate of αLP autolysis is comparable to the rate of its global unfolding, indicating that transient unfolding motions leading to proteolytic degradation have been suppressed. αLP has been so successfully optimized that it is vulnerable to degradation only after it completely unfolds, which occurs on an extremely slow time scale. In addition, the barrier to global unfolding has been maximized near the optimal growth temperature of the bacterial host.

There is a price for kinetic stability, however. The evolution of a large barrier to unfolding and a highly rigid native state through the incorporation of glycines and other changes has, as a consequence, created an even larger barrier to folding, and thermodynamically destabilized the native state of αLP. Nature's solution has been the co-evolution of a transient pro region to promote folding by both reducing the folding barrier and stabilizing the native state.

While it is expected that the general principle of longevity through kinetic stability will be shared by the majority of extracellular bacterial proteases and numerous eukaryotic proteases, the precise details of barrier height and degree of thermodynamic destabilization of the native state are likely to vary. αLP, with its large pro region and metastable native state may be an extreme example. Current studies in the lab investigating the folding of homologues of αLP, pioneered by Erin Cunningham, will be vital in testing the generality and defining the details of the continuum of kinetic stability.

References

Alonso, D.O., and K.A. Dill. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry*. 30:5974-5985.

Anderson, D.E., R.J. Peters, B. Wilk, and D. Agard. 1999. alpha-lytic protease precursor: characterization of a structured folding intermediate. *Biochemistry*. 38:4728-4735.

Arakawa, T., and S.N. Timasheff. 1982. Stabilization of protein structure by sugars. *Biochemistry*. 21:6536-6544.

Arrington, C.B., and A.D. Robertson. 1997. Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry*. 36:8686-8691.

Aune, K.C., and C. Tanford. 1969. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25 degrees. *Biochemistry*. 8:4586-4590.

Bai, Y., T. R. Sosnick, L. Mayne, and S.W. Englander. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science*. 269:5221.

Baker, D., and D.A. Agard. 1994. Kinetics versus Thermodynamics in Protein Folding. *Biochemistry*. 33:7505-7509.

Baker, D., A.K. Shiau, and D.A. Agard. 1993. The role of pro regions in protein folding. *Curr. Opin. Cell Bio.* 5:966-970.

Baker, D., J.L. Silen, and D.A. Agard. 1992a. Protease Pro Region Required for Folding Is a Potent Inhibitor of the Mature Enzyme. *Proteins*. 12:339-344.

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

Baskakov, I., and D.W. Bolen. 1998. Forcing thermodynamically unfolded proteins to fold. *J Biol Chem.* 273:4831-4834.

Baskakov, I.V., and D.W. Bolen. 1999. The paradox between m values and deltaCp's for denaturation of ribonuclease T1 with disulfide bonds intact and broken. *Protein Sci.* 8:1314-1319.

Binnie, C., L. Liao, E. Walczyk, and L.T. Malek. 1996. Isolation and characterization of a gene encoding a chymotrypsin-like serine protease from *Streptomyces lividans* 66. *Can. J. Microbiol.* 42:284-288.

Bolen, D.W., and M.M. Santoro. 1988. Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle. *Biochemistry.* 27:8069-8074.

Brayer, G.D., L.T.J. Delbaere, and M.N.G. James. 1979. Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2 Angstroms resolution. *J. Mol. Biol.* 131:743-775.

Chamberlain, A.K., T.M. Handel, and S. Marqusee. 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol.* 3:782-787.

Chen, B.L., W.A. Baase, and J.A. Schellman. 1989. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry.* 28:691-699.

Chen, X., and C.R. Matthews. 1994. Thermodynamic properties of the transition state for the rate-limiting step in the folding of the alpha subunit of tryptophan synthase. *Biochemistry.* 33:6356-6362.

Creighton, T.E. 1993. Proteins. Structures and Molecular Properties. *2nd Edition*:4.

D'Aquino, J., J. Gomez, V. Hilser, K. Lee, L. Amzel, and E. Freire. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins*. 25:143-156.

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*. 7:394-397.

Dill, K.A., and D. Shortle. 1991. Denatured states of proteins. *Annu Rev Biochem*. 60:795-825.

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

Eder, J., and A.R. Fersht. 1995. Pro-sequence-assisted protein folding. *Mol. Microbiol*. 16:609-614.

Fink, A.L. 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des*. 3:R9-23.

Gallagher, T., G. Gilliland, L. Wang, and P. Bryan. 1995. The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'. *Structure*. 3:907-914.

Gekko, K., and S.N. Timasheff. 1981. Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry*. 20:4677-4686.

Gewirth, D.T., and P.B. Sigler. 1995. The basis for half-site specificity explored through a non-cognate steroid receptor-DNA

complex. *Nat Struct Biol*. 2:386-394.

Greene, R.F., Jr., and C.N. Pace. 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem.* 249:5388-5393.

Gupta, R., S. Yadav, and F. Ahmad. 1996. Protein stability: urea-induced versus guanidine-induced unfolding of metmyoglobin. *Biochemistry.* 35:11925-11930.

Halfon, S., and C.S. Craik. 1996. Regulation of Proteolytic Activity By Engineered Tridentate Metal Binding Loops. *Journal of the American Chemical Society.* 118:1227-1228.

Henderson, G., P. Krygsman, C.J. Liu, C.C. Davey, and L.T. Malek. 1987. Characterization and structure of genes for proteases A and B from *Streptomyces griseus. J. Bacteriol.* 169:3778-3784.

Huang, G.S., and T.G. Oas. 1995. Structure and stability of monomeric lambda repressor: NMR evidence for two-state folding. *Biochemistry.* 34:3884-3892.

Huang, G.S., and T.G. Oas. 1996. Heat and cold denatured states of monomeric lambda repressor are thermodynamically and conformationally equivalent. *Biochemistry.* 35:6173-6180.

Jackson, S.E. 1998. How do small single-domain proteins fold? *Folding and Design.* 3:R81-R91.

Jackson, S.E., and A.R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry.* 30:10436-10443.

Jacob, M., T. Schindler, J. Balbach, and F.X. Schmid. 1997. Diffusion control in an elementary protein folding reaction. *Proc Natl Acad Sci U S A.* 94:5622-5627.

Kelly, J.W. 1996. Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol.* 6:11-17.

Kuhlman, B., J.A. Boice, R. Fairman, and D.P. Raleigh. 1998. Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding. *Biochemistry.* 37:1025-1032.

Lao, G., and D.B. Wilson. 1996. Cloning, sequencing, and expression of a Thermomonospora fusca protease gene in Streptomyces lividans. *Appl Environ Microbiol.* 62:4256-4259.

Lee, J.C., and S.N. Timasheff. 1981. The stabilization of proteins by sucrose. *J Biol Chem.* 256:7193-7201.

Li, R., and C. Woodward. 1999. The hydrogen exchange core and protein folding. *Protein Sci.* 8:1571-1590.

Mace, J., E., and D.A. Agard. 1995. Kinetic and Structural Characterization of Muations of Glycine 216 in α-Lytic Protease: A New Target for Engineering Substrate Specificity. *J. Mol. Biol.* 254:720-736.

Makhatadze, G.I. 1999. Thermodynamics of Protein Interactions with Urea and Guanidinium Hydrochloride. *The Journal of Physical Chemistry B.* 103:4781-4785.

Makhatadze, G.I., and P.L. Privalov. 1992. Protein Interactions with Urea and Guanidinium Chloride. A Calorimetric Study. *Journal of Molecular Biology.* 226:291-505.

Makhatadze, G.I., and P.L. Privalov. 1995. Energetics of protein structure. *Adv Protein Chem.* 47:307-425.

Matouschek, A., and A.R. Fersht. 1991. Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol.* 202:82-112.

Matouschek, A., J.M. Matthews, C.M. Johnson, and A.R. Fersht. 1994. Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions. *Protein Eng.* 7:1089-1095.

Mayr, L.M., and F.X. Schmid. 1993. Stabilization of a protein by guanidinium chloride. *Biochemistry.* 32:7994-7998.

Miranker, A., C.V. Robinson, S.E. Radford, and C.M. Dobson. 1996. Investigation of protein folding by mass spectrometry. *Faseb J.* 10:93-101.

Myers, J.K., C.N. Pace, and J.M. Scholtz. 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding [published erratum appears in Protein Sci 1996 May;5(5):981]. *Protein Sci.* 4:2138-2148.

Neurath, H. 1984. Evolution of proteolytic enzymes. *Science.* 224:350-357.

Neurath, H., and K.A. Walsh. 1976. Role of proteolytic enzymes in biological regulation (a review). *Proc Natl Acad Sci U S A.* 73:3825-3832.

Nicholson, E.M., and J.M. Scholtz. 1996. Conformational stability of the Escherichia coli HPr protein: test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. *Biochemistry.* 35:11369-11378.

Nozaki, Y., and C. Tanford. 1970. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J Biol Chem.* 245:1648-1652.

Oliveberg, M., Y.J. Tan, and A.R. Fersht. 1995. Negative activation enthalpies in the kinetics of protein folding. *Proc Natl Acad Sci U S A.* 92:8926-8929.

Otzen, D.E., O. Kristensen, M. Proctor, and M. Oliveberg. 1999. Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry*. 38:6499-6511.

Pace, C.N. 1975. The Stability of Globular Proteins. *Critical Reviews in Biochemistry*. 3:1-43.

Pace, C.N. 1986. Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods of Enzymology*. 131:266-280.

Pace, C.N., and K.E. Vanderburg. 1979. Determining globular protein stability: guanidine hydrochloride denaturation of myoglobin. *Biochemistry*. 18:288-292.

Pappenberger, G., C. Saudan, M. Becker, A.E. Merbach, and T. Kiefhaber. 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*. 97:17-22.

Parker, M.J., M. Lorch, R.B. Sessions, and A.R. Clarke. 1998. Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures. *Biochemistry*. 37:2538-2545.

Parker, M.J., J. Spencer, and A.R. Clarke. 1995. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J Mol Biol*. 253:771-786.

Perona, J.J., and C.S. Craik. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci*. 4:337-360.

Peters, R.J., A.K. Shiau, J.L. Sohl, D.E. Anderson, G. Tang, J.L. Silen, and D.A. Agard. 1998. Pro Region C-terminus: Protease Active Site Interactions are Critical in Catalyzing the Folding of alpha-lytic Protease. *Biochemistry*. 37:12058-12067.

Plaxco, K.W., J.I. Guijarro, C.J. Morton, M. Pitkeathly, I.D. Campbell, and C.M. Dobson. 1998a. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry*. 37:2529-2537.

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

Rader, S. 1997. *Thesis*.

Richards, F.M., and T. Richmond. 1977. Solvents, interfaces and protein structure. *Ciba Found Symp*. 60:23-45.

Ropson, I.J., J.I. Gordon, and C. Frieden. 1990. Folding of a predominantly beta-structure protein: rat intestinal fatty acid binding protein. *Biochemistry*. 29:9591-9599.

Royer, W.E.J., A. Pardanani, Q.H. Gibson, E.S. Peterson, and J.M. Friedman. 1996. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc Natl Acad Sci*. 93:14526-14531.

Rupley, J.A. 1967. Susceptibility to Attack by Proteolytic Enzymes. *Methods in Enzymology*. 11:905-917.

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

Sauter, N.K., T. Mau, S.D. Rader, and D.A. Agard. 1998. Structure of alpha-lytic protease complexed with its pro region. *Nature Structural Biology*. 5:945-950.

Schindler, T., and F.X. Schmid. 1996. Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry*. 35:16833-16842.

Schirmer, T., and P.R. Evans. 1990. Structural basis of the allosteric behaviour of phosphofructokinase. *nature*. 343:140-145.

Shimoi, H., Y. Iimura, T. Obata, and M. Tadenuma. 1992. Molecular structure of *Rarobacter faecitabidus* protease I: a yeast-lytic serine protease having mannose-binding activity. *J. Biol. Chem.* 267:25189-25195.

Sidhu, S.S., G.B. Kalmar, and T.J. Borgford. 1993. Characterization of the gene encoding the glutamic-acid-specific protease of *Streptomyces griseus*. *Biochem. Cell Biol.* 71:454-461.

Sidhu, S.S., G.B. Kalmar, L.G. Willis, and T.J. Borgford. 1994. *Streptomyces griseus* protease C: a novel enzyme of the chymotrypsin superfamily. *J. Biol. Chem.* 269:20167-20171.

Sidhu, S.S., G.B. Kalmar, L.G. Willis, and T.J. Borgford. 1995. Protease evolution in *Streptomyces griseus*: discovery of a novel dimeric enzyme. *J. Biol. Chem.* 270:7594-7600.

Silen, J.L., and D.A. Agard. 1989. The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. *Nature*. 341:462-464.

Silen, J.L., D. Frank, A. Fujishige, R. Bone, and D.A. Agard. 1989. Analysis of Propro-α-Lytic Protease Expression in *Escherichia Coli* Reveals that the Pro Region is Required for Activity. *J. Bacteriol.* 171:1320-1325.

Silen, J.L., C.N. McGrath, K.R. Smith, and D.A. Agard. 1988. Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme. *Gene*. 69:237-244.

Sohl, J.L. 1997. The folding and inhibition of α-lytic protease by its pro region. *Thesis*.

Sohl, J.L., S.S. Jaswal, and D.A. Agard. 1998. Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature*. 395:817-819.

Sohl, J.L., A.K. Shiau, S.D. Rader, B. Wilk, and D.A. Agard. 1997. Pro Region Inhibition by Steric Occlusion of the α-Lytic Protease Active Site. *Biochemistry*. 36:3894-3902.

Staniforth, R.A., S.G. Burston, C.J. Smith, G.S. Jackson, I.G. Badcoe, T. Atkinson, J.J. Holbrook, and A.R. Clarke. 1993. The energetics and cooperativity of protein folding: a simple experimental analysis based upon the solvation of internal residues. *Biochemistry*. 32:3842-3851.

Tanford, C. 1964. Isothermal unfolding of globular proteins in aqueous urea solution. *Journal of American Chemical Society*. 86:2050.

Tanford, C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chemistry*. 24:1-95

Timasheff, S.N. 1995. Solvent stabilization of protein structure. *Methods Mol Biol*. 40:253-269.

Yao, M., and D.W. Bolen. 1995. How valid are denaturant-induced unfolding free energy measurements? Level of conformance to common assumptions over an extended range of ribonuclease A stability. *Biochemistry*. 34:3771-3781.

Yi, Q., M.L. Scalley, K.T. Simons, S.T. Gladwin, and D. Baker. 1997. Characterization of the free energy spectrum of peptostreptococcal protein L. *Fold Des*. 2:271-280.