THE MATURATION OF ALPHA-LYTIC PROTEASE

by

AMY FUJISHIGE

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHARMACEUTICAL CHEMISTRY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

[redacted]

Acknowledgements

The completion of my Ph.D. shimmered, a mirage, for many years. In those dry years, my family and friends nourished me in every possible way. The thanks that I give them here can only stand as a small symbol of the gratitude that I hope to show in more meaningful ways in years to come.

A lifetime of thank yous belongs to my parents. Your questioning minds and thoughtful dinner conversations first kindled my interest in science. Your faith in me brought me to graduate school. Your encouragement kept me there. Thank you Mom, for not minding when I fell asleep every time I visited during those first long years of research. Thank you Dad for always calling when my spirits were lowest, "just to say hi."

My sisters, Kim, Erica, and Lee, have served as points of sanity throughout. Lee, thanks for knowing the exact stresses of graduate life. Erica, thank you for reminding me what it means to live a normal life. Kim, my confidante since childhood, you listened and listened and listened. Thank you for so much comfort through all the trials.

The members of the "orals study group," Celia Schiffer, Julie Fruetel, and Dale Bodian were fonts of information and support throughout graduate school. Special thanks to Celia for making sure that I ate and slept during the most difficult of times.

My good friends Jackie and Oriane have always believed I could do it. Thank you . (I think I really did it!)

Evi, Patience, Maria, Susan, Margaret, and Rosie-- patient housemates over the years-- put up with all the mood swings associated with failed experiments, the infrequent but gratifying successes, and the especially awful doldrums, when nothing seemed to be happening at all (but in fact, it was).

My labmates! How can I begin to thank Karen Smith, Joy Silen, and Roger Bone, who taught me what I needed to know in order to get started. I can only thank you by working carefully, and by teaching new labmates as best I can. Over the years, there have been many other labmates, all of whom are excellent intellectual foils and many who have become close friends.

I shared many cups of coffee with Julie Sohl, John Reidhaar-Olson, Elisabeth Jaffe, Jim Mace, and Barry Wilk, and Patricia Renaut. Lots of talk. Lots of help. Lots of philosophical stuff. Lots of fun. Let's stay in touch.

The crowd at Outdoors Unlimited made it possible to get out of the lab and into the wilderness. When all else fails, nature doesn't. Special thanks to Bo, whose energy and attitude toward life are infectious, and wonderful.

The Ballet Club ! Thank you Linda, Lily, Shauna, Sandra, and Kate. Our classes kept me going-- Exercise for the body, a diversion from science, and some beauty for the soul. What a great group of friends.

Over the years I have had many mentors. Tack Kuntz gave his excellent advice on both experimental protocols and the course of professional life. Peter Walter has been immeasurably helpful simply by remaining impertubable. Linda Randall is not only a mentor but a role model, very dear to my heart. I will always remember the two weeks in her laboratory, filled as they were with excellent science, dance , theater, food, wine and conversation.

Nor will I forget the mentor who deserves most thanks for all the pages that follow: my graduate research advisor, David A. Agard. I have grown both scientifically and personally under his tutelage. And it has been my special privelege to watch both David and his lab grow.

Lastly, I thank my husband, John Michael Boggs, for his generosity and support in this last year. The work is mine; its fruit is ours.

THE MATURATION OF α-LYTIC PROTEASE

AMY FUJISHIGE
UNIVERSITY OF CALIFONIA at SAN FRANCISCO

Abstract

The biogenesis of the α-lytic protease of *L. enzymogenes* has been studied in a heterologous expression system in *E. coli.* This bacterial serine protease is homologous to the mammalian enzymes of the trypsin family, yet contains a large pro region such as is found in the subtilisin family of serine proteases. The recent discovery that the pro region is required for proper folding of the protease has brought all aspects of the maturation of the initially translated precursor (prepro-α-lytic protease) under scrutiny. Studies characterizing the proteolytic processing, localization determinants, kinetics of export, and *in vitro* folding of the precursor are described. There are four major findings. First, site directed mutagenesis of the active site reveals that the precursor is autoprocessing. Next, a series of expression systems and cell fractionation techniques demonstrate that secretion across the outer membrane is intimately associated with pro-region dependent folding. These experiments suggest that the cue for export across the outer membrane is three dimensional, rather than linear. Further characterization of export by pulse-chase experiments show that export in the native host occurs more quickly than in the heterologous system. It is suggested that the native host supplies accessory proteins that aid in the export process. Finally, the temperature-sensitivity of expression seen *in vivo* is also seen *in vitro,* and the physical basis of the temperature sensitivity is shown to be a thermal instability of the pro-region. These results highlight the fact that the pro region is a multifunctional entity, whose presence affords the prokaryotic

host cell a degree of control over the protease activity in the absence of organellar compartments. Furthermore, these data emphasize that secretion systems for the outer membrane of Gram-negative bacteria are exceedingly complicated, and present a new paradigm for membrane transport and targeting. The system used here presents an excellent opportunity for further genetic and biochemical characterization of this new paradigm.

Dennis O. Good 6/10/93

Table of Contents

Acknowledgements... II

Abstract... IV

List of Tables... VII

List of Figures... VIII

Introduction... 1

References for Introduction... 7

Chapter 1: Processing Study: SA195 mutant... 11

References for Chapter One... 34

Addendum to Chapter One... 39

Chapter 2: Localization Studies... 41

References for Chapter Two... 78

Chapter 3: Kinetics of Export... 85

References for Chapter Three... 107

Chapter 4: Purification and Refolding... 111

References for Chapter Four... 136

Chapter 5: Structure-Function Studies... 138

References for Chapter Five... 181

List of Tables

Table:		Page:
I-I	Characteristics of Some Pro Regions ...	6
5-I	Crystallographic Data of α-Lytic Protease/αLP-Inhibitor Complexes ...	147
5-II	Values of Kinetic Parameters for Wild Type and Mutant αLPs ...	152
5-III	Structural Characterization of Wild Type and Mutant αLPs ...	155
5-IV	Intermolecular Hydrogen-Bond Distances in αLP-Inhibitor Complexes ...	165

List of Figures

Figure: Page:

1.1 Construction of Expression Vectors pDBR2, pALP4, and pALP5 ... 20
1.2 Fusion sites for the phoA signal sequence,
and site of Ser→ Ala mutation ... 21
1.3 Immunoblot of α-Lytic Proteases (αLP) ... 29

2.1 Construction of Deletion and Expression vectors ... 51
2.2 Constructs, Conditions, and Characteristics ... 58
2.3A) Location of 37°C Precursor by Cell Fractionation(Immunoblot) ...60
B) Co-localization of 37°C Precursor and Outer Membrane Marker ... 60
2.4 Co-localization of SA195 Precursor and Outer Membrane Marker ... 61
2.5 Co-localization of D-pro-αLP and Outer Membrane Marker ... 62
2.6 Co-localization of Precursor and
Outer Membrane Marker at High pH ... 63
2.7 Accessibility of Precursor to Trypsin ... 66
2.8 Correlation of Activity and Secretion in Pro-Deletion Mutants ... 67
2.9A) Location of pCOMP4-SA195 ... 69
B) Native Gel of pCOMP4-SA195 ... 69
2.10 Location of the Pro Region ... 71
2.11 Model of α-Lytic Protease Maturation ... 76

3.1 Time Course of Expression of αLP ... 97
3.2 α-Lytic Protease is Secreted via the Periplasm
in *E.coli* (Pulse Chase) ... 99
3.3 Kinetics of Secretion of αLP across
the Outer Membrane of *E. coli* ... 100
3.4 Pulse chase of αLP in its Native Host, *Lysobacter enzymogenes* ... 102
3.5 Kinetics of Secretion of αLP in *Lysobacter enzymogenes* ... 103

4.1 Temperature Dependence of *in vivo* Expression ... 123
4.2 Purification and Refolding of 37°C Precursor of αLP:
A) MA7C column analysis 125
B) Superose 12 column analysis 126

Figure: Page:

4.3 Trypsinization Timing Optimum ... 127
4.4 Refolding: pH Optimum ... 128

4.5 Temperature Sensitivity of Folding (precursor and pro+mature) ...130
4.6 Thermal Melt of the Pro Region Measured by Circular Dichroism...131

5.1 Schematic of the Interactions between αLP and Inhibitors ... 148
5.2 Correlation of Ki and kcat/Km for Wild Type and Mutant αLPs ... 160
5.3 Superimposition of Wild Type αLP Specificity Pocket Complexed with each of Five Inhibitors ... 161
5.4 Superimposition of MA192 Mutant Specificity Pocket Complexed with each of Five Inhibitors ... 162
5.5 Superimposition of MA213 Mutant Specificity Pocket Complexed with each of Five Inhibitors ... 163

Introduction

α-Lytic protease is one of several enzymes secreted by the Gram negative soil bacterium, *Lysobacter enzymogenes* (Whitaker, 1970). Because it is a serine protease of the trypsin family, it has been used extensively as a model system for understanding serine protease mechanism and protein-ligand specificity. It is uniquely suited for this purpose in two ways. First, the histidine of the catalytic triad is the sole histidine in α-lytic protease. The corresponding unique nuclear magnetic resonance signal from this key residue has allowed researchers to probe the enzyme mechanism (Bachovchin et al., 1981). Secondly, the protease has been crystallized and the structure refined to 1.7 Å (Fujinaga et al., 1985). This structure reveals that the ligand-binding pocket is lined with the side chains of three amino acids. This has allowed us to probe the structural basis for specificity by employing site directed mutagenesis in conjunction with kinetic and structural characterization of the mutant enzymes (Bone et al., 1987; Bone et al.,1989a,b; Bone et al., 1991).

Perhaps the most intriguing aspect of this enzyme, however, is the maturation of the precursor polypeptide to the mature, active protease. Precursor proteins offer the cell a level of regulation that can be temporally and sometimes spatially controlled. Their maturation may include glycosylation, conformational changes, or proteolytic processing. In the case of α-lytic protease and several other proteins, activity and stability are regulated by the presence and removal of a pro region. Pro regions are large polypeptides that are excised from precursors to produce the mature protein. They may occur at either terminus or internally. How do these pro regions regulate protease activity? An increasing number of pro regions appear to mediate folding, and in some cases secretion, of their contiguous protein

(Table I-I). The ability of the pro region to lower the rate limiting barrier in folding and the fact that it is not required for activity thereafter allows α-lytic protease to achieve a highly stable final conformation (Baker et al., 1992b). In addition, the pro region may bind the final product in such a way as to inhibit its activity (Baker et al., 1992a; Ohta et al., 1991). The emerging data on various proenzymes appears to define a new functional class of precursor.

Among precursors, it is necessary to distinguish the pro region from both the "pre" region and the small zymogen activation peptides. A pre region is a small hydrophobic peptide, or signal sequence, that mediates targeting to or through any of several cytoplasmic membranes (Blobel and Dobberstein, 1975). Recent information indicates that signal sequences also interact with the nascent protein to slow down folding, possibly in order to allow more time for chaperones (see below) to bind (Liu et al., 1989). By contrast, pro regions can be highly charged, and are usually longer than a typical signal sequence. Pro regions occur both with and without an attached pre region. Targeting across a cytoplasmic membrane is usually signal sequence mediated, while targeting to distal destinations may involve the pro region (Pohlner et al., 1987; Terada et al., 1990; Fujishige et al., 1992).

Zymogens are a particular kind of precursor form of an enzyme. The zymogen includes an "activation" peptide whose presence greatly diminishes or abolishes activity. Removal of the peptide allows full expression of activity. In the well characterized case of trypsinogen, cleavage of the activation peptide (the first 6 amino acids of the precursor) creates a new N-terminus (Kunitz, 1936; Maroux et al., 1971). This positively charged group is then resituated to make a salt bridge which, with other conformational rearrangements, is necessary to the mechanism of trypsin and other serine proteases (Stroud et al., 1977). Interestingly, many large pro regions appear to

inhibit their contiguous proteins. This interaction can be very tight and can persist even after the cleavage of the pro region from the companion protein (Ohta et al., 1991; Baker et al., 1992a). Unlike an activation peptide, the pro region of α-lytic protease binds the native state of the enzyme. Thus, both pro regions and activation peptides can preclude activity, but they use distinctly different mechanisms. Most importantly, pro regions appear to be required for folding while the activation peptide is not. For example, trypsin can be expressed in active form without its activation peptide (Vasquez et al., 1989), but α-lytic protease and subtilisin cannot be produced in vivo in active form except in the presence of their pro regions (Silen et al., 1989; Ikemura et al., 1987).

Yet another distinction can and should be made between pro regions and molecular chaperones. Like pro regions, the molecular chaperones aid in the folding, assembly, and/or secretion of proteins. Unlike pro regions, chaperones are not at any time covalently bound to their substrates. Molecular chaperones are promiscuous with regard to their substrates, whereas pro regions are specific. This fact may explain why chaperones generally utilize ATP, whereas pro regions appear to take advantage of binding energy and proteolytic degradation. Furthermore, chaperonins appear to function by preventing off-pathway reactions, such as aggregation, or by retaining the protein in an assembly- or secretion-competent state (for review of chaperones, see Gething and Sambrook, 1992). By contrast, the α-lytic protease pro region appears to promote folding by binding to a high energy, rate limiting transition state in the folding pathway (Baker et al., 1992b). While chaperones appear to be reusable, pro regions may be compared to single turnover enzymes.

As described in previous publications and in this dissertation, the maturation of α-lytic protease is quite different from that of the well-studied maturation of its mammalian counterpart, trypsin. α-Lytic protease is responsible for its own processing, as evidenced by the accumulation of uncleaved proenzyme on expression of the active site mutant, Ser 195-> Ala 195 (SA195; Silen et al., 1989). The N-terminal pro region of α-lytic protease is much larger than trypsin's activation peptide, and has now been shown to possess much wider functions than does the activation peptide. The pro region of α-lytic is required for proper folding and secretion of the enzyme, and for its inhibition. Heterologous expression in *E. coli* of the preproenzyme allows the accumulation of accurately processed, active enzyme in the medium, whereas expression of the DNA that encodes only the mature enzyme (lacking the sequence for the large pro region, Δ-pro-α-lytic protease) gives rise to inactive α-lytic protein that accumulates in the outer membrane (Silen et al., 1989; Fujishige et al., 1992). By contrast, active trypsin can be obtained by expression in *E. coli* of either the zymogen or the mature form (C.S. Craik, pers. comm.) Moreover, separate pro region and protease region polypeptides can be combined to produce folded, active protease either *in vivo* (Silen and Agard, 1989) or *invitro* (Baker et al., 1992a). *In vivo*, it appears that only the correctly folded or folding competent form of α-lytic protease is correctly targeted to the extracellular medium. Misfolded Δ-pro-α-lytic and uncleaved precursor forms accumulate in the outer membrane. Even cleaved from the protease, the pro region serves a second function, which is that of a potent inhibitor to the protease (Baker et al., 1992b). Consistent with its subnanomolar binding constant, the pro region has not been observed to dissociate as an intact polypeptide. Rather, the cleaved but noncovalently bound pro region appears to dissociate from the protease only after partial

degradation by other proteases or by α-lytic itself. Apparently, the pro region-dependent folding and inhibition are a distinct form of protective regulation of the host. In addition, the combination of pro region-assisted folding and pro region-independent final function allows the formation of a highly stable protease.

In summary, despite their many similarities to molecular chaperones, activation peptides, and signal sequences, pro regions are mechanistically distinct from any of these. Results from the study of the maturation of α-lytic protease may serve as a model for pro region mediated functions. In addition, these studies facilitate our increasingly detailed structure-function studies on α-lytic protease, and provide a basis for further pursuits in the areas of extracellular protein secretion in Gram-negative bacteria.

Table I-I: Examples of pro regions and their tested biological roles

Protein	mature	pro	Host	probable functions
α-lytic protease	198 a.a.	166 a.a.	*L. enzymogenes*	folding, inhibition[a]
subtilisin E	275	77	*B. subtilis*	folding, inhibition[b]
SGPA, SGPB	181,185	76,78	*S. griseus*	folding[c]
aqualysin I	281	113, 105	*T. aquaticus*	folding[d]
SSP	381	637	*S. marcescens*	secretion[e]
AEP	297	124	*Y. lipolytica*	folding, targeting[f]
CpY	329	79	*S. cerevisiae*	folding, inhibition[g]
papain	~220	~100	*C. papaya*	folding[h]
TGF-β1	112	250	*H. sapiens*	dimerization, targeting[i]
activin A	116	290	*H. sapiens*	dimerization, targeting[i]

a.a. = amino acids

SGPA= Streptomyces griseus protease A. SSP = Serratia marcescens Serine Protease. AEP = Alkaline Extracellular Protease. CpY = Carboxypeptidase Y. TGF-β1 = Transforming Growth Factor-β 1

a) Silen et al., 1989; Silen and Agard, 1989; Baker et al., 1992a
b) Ikemura et al., 1987; Zhu et al., 1989; Ohta et al., 1991
c) Henderson et al., 1987; J. Couto, pers. comm.
d) Terada et al., 1990
e) Miyazaki et al., 1989
f) Fabre et al., 1992
g) Winther and Sørensen, 1991
h) Vernet et al., 1990
i) Gray and Mason, 1990

References for Introduction

Bachovchin, W. W., R. Kaiser, J. H. Richards, and J. D. Roberts. 1981. Catalytic mechanism of serine proteases: reexamination of the pH dependence of the histidyl 1 J_{3C2-H} coupling constant in the catalytic triad of α-lytic protease. Proc. Natl. Acad. Sci. USA **78**:7323-7326.

Baker, D., Silen, J.L., and D.A. Agard. 1992a. A protease pro region required for folding is a potent inhibitor of the mature enzyme.*Proteins* 12:339-344

Baker, D., Sohl, J.L., and D.A. Agard. 1992b. A protein-folding reaction under kinetic control. *Nature.* 365:263-265

Blobel, G. and B.Dobberstein. 1975. Transfer of proteins across membranes. *J. Cell Biol.* 67:835-851

Bone, R.B., Frank, D., Kettner, C.A., and D.A.Agard. 1989a. Structural analysis of specificity: α-lytic protease complexes with analogues of reaction intermediates. *Biochemistry* 28:7600-7609

Bone, R.B., Fujishige, A., Kettner, C.A. and D.A. Agard. 1991. Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry:* 30:10388-10398

Bone, R.B., Shenvi, A.B., Kettner, C.A., and D.A.Agard. 1987. Serine protease mechanism: Structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. *Biochemistry* 26:7609-7614

Bone, R.B., Silen, J.L., and D.A.Agard. 1989b. Structural plasticity broadens the specificity of an engineered protease. *Nature* 339:191-195

Fabre, E., Tharaud, C. and C. Gaillardin. 1992. Intracellular transit of a yeast protease is rescued by trans-complementation with its prodomain. *J. Biol. Chem.* 267:15049-15055

Fujinaga, M., L.T.J. Delbaere, G. D. Brayer, and M. N. G. James. 1985. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. **183**:479-502.

Fujishige, A., Smith, K.R., Silen, J.L., and D.A. Agard. 1992. Correct folding of α-lytic protease is required for its extracellular secretion from *Escherichia coli. J. Cell Biol.* 118:33-42

Gething, M.-J., and J. Sambrook. 1992. Protein folding in the cell. *Nature* 355:33-45

Gray, A.M., and A. J. Mason. 1990. Requirement for activin A and transforming growth factor-β1 pro-regions in homodimer assembly. *Science* 247:1328-1330

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

Ikemura, H. H. Takagi, and M. Inouye. 1987. Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli. J. Biol. Chem.* 262:7859-7864.

Kunitz, M., and J.H. Northrup. 1936. Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *J. Gen. Physiol.* 19:991-1007

Liu G; Topping TB; Randall LL. 1989. Physiological role during export for the retardation of folding by the leader peptide of

maltose-binding protein. *P.N.A.S. (U.S.A.)* 86:9213-7.

Maroux S; Baratti J; Desnuelle P. 1971. Purification and specificity of porcine enterokinase. *Journal of Biological Chemistry*

Miyazaki, H., Yanagida, N., Horinouchi, S., and T.Beppu. 1989. Characterization of the precursor of *Serratia marcescens* serine protease and COOH-terminal processing of the precursor during its excretion through the outer membrane of *Escherichia coli. J. Bact.* 171:6566-6572

Ohta Y; Hojo H; Aimoto S; Kobayashi T; Zhu X; Jordan F; Inouye M. 1991. Pro-peptide as an intramolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide [corrected] [published erratum appears in Mol Microbiol 1991 Dec;5(12):3090]. *Molecular Microbiology* 5:1507-10.

Pohlner, J., Halter, R., Beyreuther, K., and T.F.Meyer. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* 325:458-462

Robertus JD; Alden RA; Birktoft JJ; Kraut J; Powers JC; Wilcox, P.E. 1972. An x-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry* 11:2439-49.

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., Frank, D., Fujishige, A., Bone, R., and D.A. Agard. 1989. Analysis of prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro region is required for activity. *J. Bact.* 171:1320-1325

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

Stroud, R. M., Kossiakoff, A.A., and J.L. Chambers. 1977. Mechanisms of zymogen activation. *Ann. Rev. Biophys. Bioeng.* 6:177-193

Terada, I., Kwon, S., Miyata, Y., Matsuzawa, H., and T.Ohta. 1990. Unique precursor structure of an extracellular protease, aqualysin I, with NH_2- and COOH-terminal pro-sequences and its processing in *Escherichia coli.* *J. Biol. Chem.* 265: 6576-6581

Vasquez, J.R., Evnin, L.B., Higaki, J.N., and C.S. Craik. 1989. An expression system for trypsin. *J. Cell. Biochem.* 39:265-276

Vernet, T. Tessier, D.C., Richardson, C., Laliberte, F., Khouri, H.E., Bell, A.W., Storer, A.C., and D.Y.Thomas. 1990. Secretion of functional papain precursor from insect cells. *J. Biol. Chem.* 265:16661-16666

Whitaker, D.R. 1970. The α -lytic protease of a myxobacterium. *In* Meth. Enz. 19:599-613

Winther, J. R., and P. Sørensen. 1991. Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. *Proc. Nat. Acad. Sci. (U.S.A.)* 88:9330-9334

Zhu, X., Ohta, Y., Jordan, F., and M. Inouye. 1989. Pro-sequence of subtilisin can guide the refolding of tdenatured subtilisin in an intermolecular process. *Nature* 339:483-484

Chapter 1: **Processing Studies**

α-lytic protease is translated as a preproenzyme. The pre region is a typical signal sequence. The pro region is transiently required during the folding process in order to achieve the active conformation of the catalytic region. Neither the pre nor the pro region is present in the final product, the mature protease. Processing of signal sequences has been studied extensively in other laboratories and appears to occur by a general mechanism. The processing of pro regions is less studied. Thus far, the characteristics of such processing appear to be system-specific. I undertook to examine this important aspect of the maturation of α-lytic protease by using a combination of molecular biology, biochemistry, and enzymology.

A major question confronting us was whether α–lytic protease is itself responsible for the proteolytic cleavage that separates the pro region from the catalytic region. In order to address this question, I mutated the active site serine residue to an alanine. The removal of the nucleophilic oxygen of the serine disables a serine protease. Therefore, if α-lytic protease is self-processing with regard to its pro region, the catalytic knockout should result in the accumulation of the proenzyme. If, on the other hand, the pro region is cleaved from the catalytic region by another protease, then the maturation process should be undisturbed and result in the accumulation of mature size (albeit inactive) protease.

These processing studies were published as part of a larger body of work from the Agard lab. This chapter is comprised of that paper (Journal of Bacteriology, 1989, Volume171, pages 1320-1325) following which I have added my comments concerning our understanding of processing.

Analysis of prepro-α-lytic protease expression in *E. coli* reveals that the pro region is required for activity.

Joy L. Silen, Dan Frank, Amy Fujishige, Roger Bone, and David A. Agard

Department of Biochemistry & Biophysics and
The Howard Hughes Medical Institute,
University of California at San Francisco,
San Francisco, CA 94143-0448 (U.S.A.) Tel. (415)-476-5143

Abstract

The α-lytic protease of *Lysobacter enzymogenes* has been successfully expressed in *Escherichia coli* by fusing the promoter and signal sequence of the *E. coli phoA* gene to the proenzyme portion of the α-lytic gene. Following induction, active enzyme was found both within cells and in the extracellular medium, where it slowly accumulated to high levels. Use of a similar gene fusion to express the protease domain alone produced inactive enzyme, indicating that the large amino terminal pro region is necessary for activity. The implications for protein folding are discussed. Furthermore, inactivation of the protease by mutation of the catalytic serine residue resulted in the production of a higher molecular weight form of α-lytic protease, suggesting that the enzyme is self-processing in *E. coli*.

Introduction

α-Lytic protease is one of a battery of extracellular enzymes secreted by the Gram-negative bacterium *Lysobacter enzymogenes* to lyse and degrade soil microorganisms. By virtue of its well-studied Asp-His-Ser catalytic triad mechanism (Bachovchin and Roberts, 1978; Bachovchin et al., 1981), its degree of structural homology to mammalian serine proteases (Brayer et al., 1979), and its amenity to NMR studies (Bachovchin and Roberts, 1978; Bachovchin et al., 1981; Robillard and Shulman, 1974), α-lytic protease is an ideal candidate for site-specific mutagenesis studies of substrate specificity and structure-function relationships (Bone et al., 1987). We recently reported the cloning and sequence analysis of the α-lytic protease gene from *L. enzymogenes* (Silen et al., 1988). The nucleotide sequence contained a large open reading frame 5' of the mature enzyme's coding sequence, and we proposed that the additional 199 amino acids comprised a 37 amino acid signal sequence ("pre") and a 165 amino acid "pro" region . Recent studies with secreted proteases from both Gram-positive and Gram-negative bacteria, including several *Bacillus* species (Shimada et al., 1985; Vasantha et al, 1984; Stahl and Ferrari, 1984; Takagi et al., 1985; Wells et al., 1983; Yang et al., 1984), *Neisseria gonorrhoeae* (Pohlner et al., 1987), *Streptomyces griseus* (Henderson et al., 1987) and *Serratia marcescens* (Yanagida et al., 1986) have shown that all of these bacterial proteases are synthesized as precursors, although the pro region varies in its amino- or carboxyl-terminal location. Recently the 77 amino acid amino-terminal pro region of *Bacillus subtilis* subtilisin E, has been shown to be necessary for the production of active protease, suggesting a critical role for the pro region in folding (Ikemura et al., 1987). In this paper we show that the pro region of α-lytic protease has a similar function.

We have subcloned regions of the α-lytic protease gene behind the promoter and signal sequence of the *Escherichia coli (E. coli)* gene *pho*A, allowing production of α-lytic protease in *E. coli* under conditions of phosphate depletion (Inouye et al., 1981). In addition to confirming the requirement for the pro region, this approach has provided evidence that expression of active protein is temperature-sensitive in *E. coli*, and that α-lytic protease has the ability to proteolytically process itself. The implications of these results for protein folding are discussed.

Materials and Methods

Enzymes-- All restriction enzymes were purchased from New England Biolabs with the exception of *Sma*I, which was purchased from Boehringer Mannheim.T4 polynucleotide kinase was purchased from P-L Pharmacia. Calf intestinal phosphatase was purchased from Boehringer Mannheim and column purified (Efstratiadis et al., 1977). The Klenow fragment of DNA polymerase I was purchased from Amersham. All enzymes were used as recommended by the manufacturers.

Strains, media and expression-- DG 98, an F' carrying strain used for M13 phage growth, was previously referenced (Silen et al., 1988). MH1 (araD139, Δlac X74, galU, galK, hsr⁻rpsL) was used as the host for expression of α-lytic protease (Hall et al., 1984). For M13 phage growth, YT medium was used (Miller, 1972). LB medium containing 50 ug/ml ampicillin was used to maintain plasmids derived from pBR329.

For *pho*A directed expression of α-lytic protease we used MOPS medium (Neidhardt et al., 1974), supplemented with 0.4% glucose, 0.15% vitamin free casamino acids (Difco) and $KHPO_4$ or yeast extract to provide a controlled level of phosphate. Typically, after growth overnight from a single colony in 5 mls of high-phosphate MOPS (50mM KHPO4, pH 7.4, 50 ug/ml ampicillin) cells were pelleted, washed three times, and resuspended in phosphate-free MOPS medium. The cells were then diluted 1 to 50 into MOPS supplemented with 0.15% yeast extract [0.15% yeast extract was determined to provide 0.63 mM phosphate (total)] , which allowed the cells to reach a high density (A_{600} = 3) before the phosphate was depleted. Expression was monitored either by alkaline phosphatase induction or by the production of α-lytic protease (see enzyme assays). Typically, maximal production of α-lytic protease was

achieved after 5 days of induction at 24°C. For plate screening, colonies were transferred to low (0.1mM) phosphate plates for induction at room temperature, then onto milk plates (LB agar with 0.5% skim milk powder) (37), where a region of clearing indicated protease activity.

Plasmid constructions-- To be consistent with the published literature on α-lytic protease we have numbered the mature amino acid sequence 1-198* and have assigned negative numbers to the prepro sequence. The putative signal sequence, therefore, is numbered -199 to -166, and the pro region -165 to -1 (Silen et al., 1988).

M13mp18 and mp19 replicative forms were purchased from Boehringer Mannheim. pHI28, which contains the *pho*A promoter and signal sequence, followed by a polylinker region, was the generous gift of H. Inouye and S. Michaelis. When necessary, cohesive ends were filled-in with Klenow fragment in nick translation buffer. All restriction fragments were isolated in 0.5- 2.0% low-gelling temperature agarose gels, and ligations were performed in diluted agarose (Struhl, 1985). Transformation was accomplished by the $MgCl_2$, $CaCl_2$ method (Morrison, 1989). Plasmids and replicative forms were isolated by alkaline lysis (Birnboim and Doly, 1979). Restriction enzyme digest analyses were resolved on 0.7- 2.0% agarose gels.

pDA27, an M13mp19 derivative that contains the *pho*A promoter and signal sequence, was constructed by inserting the *Pst*I/*Xma*I fragment from pHI28 into the *Pst*I/*Xma*I sites of M13mp18. pDBR2, our basic expression vector, was made by inserting the *Eco*RI/*Sph*I fragment of pDA27 into the corresponding sites of pBR329 (Fig. 1.1a).

pALP4, the construct that placed the α-lytic proenzyme under the control of the *pho*A promoter and behind the *pho*A signal sequence, was made by inserting a 1.3 kb, filled-in EagI fragment from pDA13 (M13mp18

containing the α-lytic gene on a 1741bp *Eco*RI fragment; Silen et al, 1988) into *Sma*I cut pDBR2 (Fig. 1.1b, Fig.1. 2). Insertion in the correct orientation was ascertained first by screening for enzymatic activity, then by restriction analysis.

Oligonucleotide mutagenesis– Purified oligonucleotides were provided by the Biological Resource Center (UCSF), with the exception of the *Xho*I mutation oligomer, which was provided by the Howard Hughes Medical Institute (UCSF). Site specific mutagenesis was performed by the double primer method of Norris et al. (1983) with minor modifications. In all cases, the P-L M13 sequencing primer (17- mer) was used as the second (nonmutagenic) primer. After extension with Klenow fragment in the presence of ligase, the hybrid molecules were either digested with restriction enzymes and the appropriate fragment gel purified and recloned into a new M13 vector (Norrander et al., 1983), or, alternatively, were directly transformed into DG98. Screening of transformants was done either by gel analysis of replicative forms (for the creation of restriction enzyme sites), or by hybridization to plaque lifts with ^{32}P-labelled oligonucleotide ([^{32}P]-γ- ATP Amersham, > 3000 Ci/mmol). The presence of mutations was confirmed by the chain-termination sequencing method of Sanger et al. (1977).

pDA16, a *Nar*I fragment in M13mp19 that contains the mature coding sequence (in the positive orientation relative to β-galactosidase transcription) was mutated by site specific mutagenesis to contain two unique restriction sites: *Mlu*I (base 1248 of the EcoRI fragment), using GCGGCCCATGCACGCGTTGC CCT, and *Xho*I (base 1375), using TCGAACAGGCTCGAGCGCTGCGAGG to make pDA16XM. The *Bal*I/*Sma*I fragment from the pDA16XM replicative form was substituted for the corresponding *Bal*I fragment in pALP4 to make pALP5 (Fig.

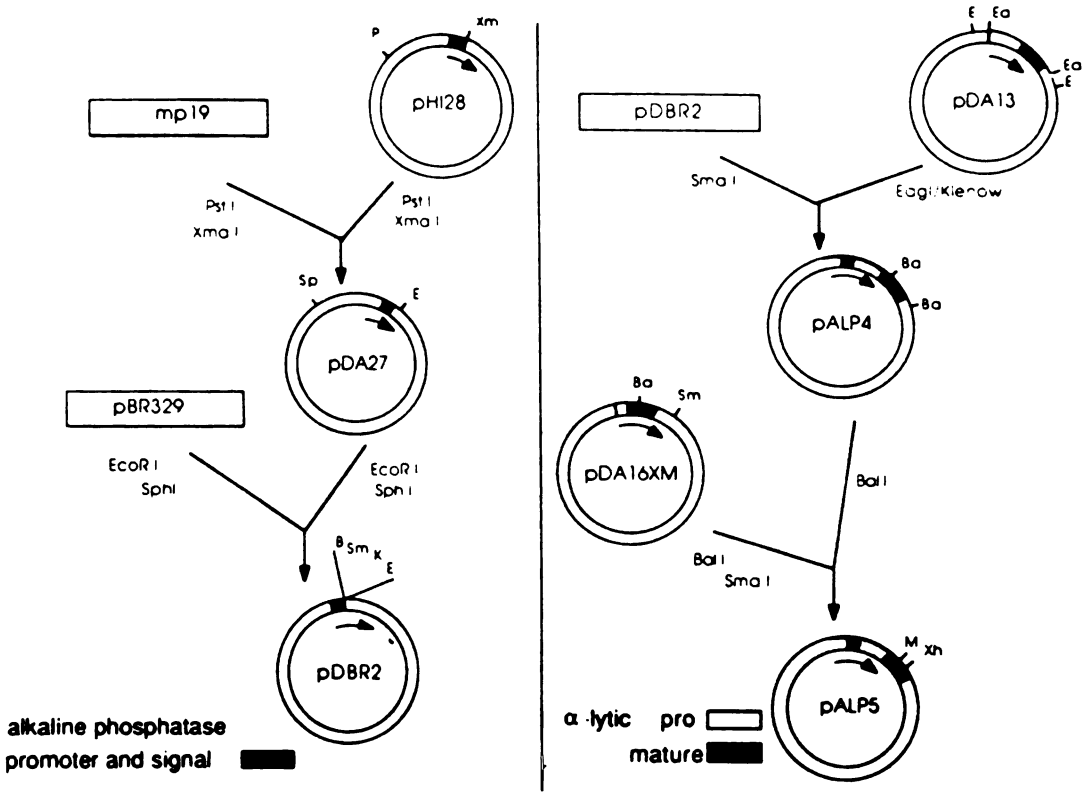


Fig. 1.1. The construction of α-lytic protease expression vectors using the *pho*A promoter and signal sequence of pHI28. a)The construction of pDBR2. b) the construction of pALP4 and pALP5. Ba, *Bal*I; B, *Bam*HI; Ea, *Eag*I; E, *Eco*RI; K, *Kpn*I; M, *Mlu*I; P, *Pst*I; Sm, *Sma*I; Sp, *Sph*I; Xm, *Xma*I; Xh, *Xho*I. For details see materials and methods.

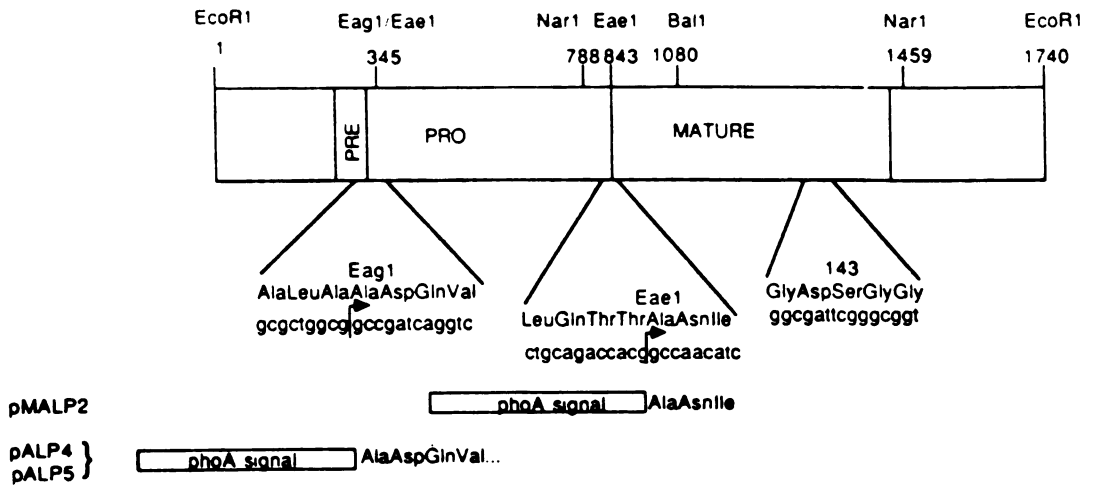


Fig.1. 2. Fusion sites of the *pho*A signal sequence to α-lytic protease in pMALP2, pALP4, and pALP5, and the site of the catalytic serine [Ser 143 (195*)]which was substituted with alanine to make an inactive protease.

pMALP2, the construct that allows expression of the protease domain (amino acids 1-198) alone, and in which the protease is directly fused to the *pho*A signal sequence, was made by oligonucleotide-directed deletion. First the mature coding region was placed 3' of the *pho*A promoter and signal sequence by inserting a *Kpn*I/filled-in *Hin*dIII fragment from pDA15 (the *Nar*I fragment of α-lytic protease in M13mp19 in the minus orientation) into the *Kpn*I/ filled-in *Eco*RI site of pDBR2 polylinker to make pMALP1. To produce single-stranded DNA for mutagenesis, the *Sph*I fragment of this construct was cloned into the M13mp18 *Sph*I site and checked by restriction analysis for the proper orientation. To eliminate the intervening 11 amino acids contributed by the M13 polylinker and the 18 amino acids from the α-lytic protease pro region, we employed an oligonucleotide-directed deletion using the 30 mer 5'-CCTGTCACAAAGCC/GCCAACATCGTCGGC -3' (the slash indicates the fusion site, see oligonucleotide mutagenesis). The *Sph*I fragment from the mutated plasmid (which included the *pho*A signal sequence /mature α-lytic protease fusion) was transferred into *Sph*I- cut pMALP1, to make pMALP2 (Fig. 1.2).

The serine 195* to alanine mutation (pALP5 SA195) was accomplished by site-specific mutagenesis of pDA16XM single-stranded DNA, using the 25mer CGAACCGCCAGCATCGCCGCGGCCC. pALP5 DNA was cut with *Mlu*I and *Xho*I and the mutated fragment from the pDA16XM replicative form was substituted .

Protein analysis-- 15% polyacrylamide SDS gels were run according to Laemmli(1970). Western blots (Burnette, 1981) were probed with affinity-purified rabbit anti-α-lytic protease antibody followed by ^{125}I protein A (ICN, >300uCi /ug). Antibody immunoserum was prepared by Babco, Berkeley, CA, and was purified using Affigel-10 (Biorad) derivatized (Naito and Ueda, 1981) with native α- lytic protease. Gel samples were prepared by pelleting cells at 12,000 x g in a Beckman microfuge and boiling in sample buffer, or for culture

supernatant, dialysed against water then combined with sample buffer and boiled.

Protein purification and enzyme assays– Native α -lytic protease was purified from culture filtrates of *Lysobacter enzymogenes* 495 (Hunkapiller et al, 1973; Whitaker, 1970) and migrated as a single band when electrophoresed on a native polyacrylamide gel at low pH (Hames and Rickwood, 1981). The enzyme was assayed using succinyl-Ala-Pro-Ala-*p*-nitroanilide (Peninsula Chemicals) as the substrate by monitoring p-nitroaniline production at 410nm (ΔE = 8860 M cm). Alkaline phosphatase production was monitored at 420nm as described by Brickman and Beckwith (Brickman and Beckwith, 1975) using *p*-nitro-phenyl phosphate as the substrate.

Results

Expression of active α-lytic protease-- We previously reported the cloning of the entire α-lytic protease gene, including regulatory elements, contained on a single *Eco*RI fragment in M13mp18 (pDA13; Silen et al., 1988). We were unable to detect expression of α-lytic protease in *E. coli* strains containing this gene either by enzymatic assay or by immunological screening. The lack of activity was attributed to the poor agreement of the

-10,-35 promoter sequence and ribosome binding site sequences relative to *E. coli* consensus sequences (Silen et al., 1988), although it was also possible that the promoter required *L. enzymogenes*-specific expression proteins not present in the phylogenetically distant *E. coli*.. For our α-lytic protease structure-function studies it was necessary to develop a suitable *E. coli* expression system. As α-lytic protease is likely to require an oxidative environment for the proper formation of its three disulfide bonds (Olson et al.,1970), we substituted an *E. coli* signal sequence (in addition to the promoter and rbs region) for the corresponding *L. enzymogenes* sequences to ensure transport to the periplasm. Studies with subtilisin E (Ikemura et al., 1987), among other bacterial proteases, have indicated that the long amino-terminal pro region was required for expression of active enzyme. As the nucleotide sequence of α-lytic protease suggested a preproenzyme structure, we fused the *pho*A promoter and signal sequence to the α-lytic pro region at the point we judged, by amino acid composition, to be the end of the α- lytic signal sequence (filled in EagI site, base 342; see Fig. 1.2).

The resulting construct, when transformed into MH1, conferred phosphate starvation-inducible protease activity. Surprisingly, we found that this activity was temperature sensitive and was not detectable at temperatures

above 30°C. After induction in liquid culture at 24°C, the activity was detectable in both the cells and the culture medium; with increasing time the majority was found in the medium (activity corresponding to 6 mg/l after 5 days). This distribution was confirmed by Western blot analysis (Fig. 1.3), where antigenically cross-reacting bands from both the cell pellet and the supernatant appeared at the same molecular weight as the purified *L. enzymogenes* enzyme. No higher molecular weight form was detected, indicating that processing of the preproenzyme had already occurred. Purification of the enzyme from liquid culture medium established that the protease expressed in *E. coli* had identical kinetic parameters to the enzyme produced by *L. enzymogenes*. The values of k_{cat} and K_m were 79 s^{-1} and 5.7mM for the *L. enzymogenes* enzyme and 73 s^{-1} and 5.6 mM for the enzyme produced in *E. coli*.. Amino-terminal sequence analysis of the purified *E.coli* product revealed that cleavage from the pro region occurs at the same site as in the *L. enzymogenes* enzyme (data not shown).

When foreign proteins expressed in *E. coli* are detected in the medium, there is always the question of whether the transport is specific, as in the case of IgA protease of *Neisseria gonorrhoeae* (Pohlner et al., 1987), or due to leakiness of the outer membrane, usually indicated by the presence of periplasmic enzymes in the culture medium. Unfortunately, in our current system, expression is achieved upon the cessation of growth (Michaelis et al., 1983), and protease production is measured over days, not hours. The length of the expression time makes the question of selective transport from the periplasm more difficult to address, as one must compensate for cell death and lysis in addition to the effects of the expressed protein. However, when parallel cultures of MH1/pALP5 and MH1/pBR329 were compared after only 12 hours of induction, the pALP5 culture had significant levels of β-lactamase,

alkaline phosphatase, and cyclic phosphodiesterase (all periplasmic enzymes) in the medium, while the pBR329 culture showed no such leakage (data not shown). This apparent loss of integrity of the outer membrane suggests transport is non-specific.

To facilitate the generation of binding pocket mutations we designed a derivative of pALP4, pALP5, creating 2 unique restriction sites: *Mlu*I (base 1248 of the *Eco*RI fragment) and *Xho*I (base 1375). These sites flanked the coding sequence for amino acids 136 to 177, and did not change the amino acid sequence. Cell extracts of MH1 cells expressing protein from pALP5 gave identical results to those found with pALP4 (Fig. 1.3), as judged by Western analysis and enzymatic assay. This construct was used in all expression studies unless otherwise stated.

The role of the precursor in α-lytic expression in E. coli-- To assess the role of the pro region in the expression of functional α-lytic protease, we examined the activity of the protease domain expressed alone. Using an oligonucleotide-directed deletion, we made a plasmid containing a direct fusion of the final codon of the *pho*A signal sequence (Ala) to the first amino acid of mature α-lytic protease (Ala 1) (pMALP2, see Fig. 1.2). When MH1 cells carrying this construct were grown under conditions of phosphate starvation, we were unable to detect activity in either the supernatant or in cell pellets. Western analysis revealed an immunologically cross-reacting species that migrated identically to purified *L. enzymogenes* α-lytic protease on both SDS-PAGE gels (Fig. 1.3) and on non-reducing native gels. Like the active enzyme, the inactive protease domain appeared to accumulate with time, reaching levels (approximately 3 mg/l) much greater than that required to assay even fractionally-active enzyme (data not shown). However, unlike the active construct, the protease domain remained associated with the cell pellet;

proportionally little was found in the medium. The gel migration behavior indicated that the *pho*A signal sequence was cleaved, suggesting a periplasmic localization. The lack of activity from this construct indicated that production of active α-lytic protease in *E. coli* requires the pro region.

An interesting feature of the production of active protease is its temperature-sensitivity. Expression at 37°C results in no detectable activity, and reveals no immunologically cross-reacting protein at the expected size of the mature form (20 kd). However, a cross-reacting protein of about 40 kd, the expected size of the proenzyme is produced (the presence or absence of the signal sequence was not established) (Fig. 1.3). As was the case with pMALP2 expression, most of the protein appeared to remain cell-associated and presumably accumulated in the periplasm. The absence of the processed form suggested that the enzymatic activity normally associated with α-lytic protease might also be responsible for the processing of the pro-enzyme and the release of the mature form.

To directly test the role of the protease in the processing of the proenzyme, we made a mutation that changed the codon for the catalytic serine 195* (amino acid 143) to an alanine, rendering the enzyme inactive. Expression of the protein from this plasmid at 24°C gave a result similar to that seen for expression at higher temperatures: production of an inactive, higher molecular weight form of the predicted pro-enzyme size (41Kd, see Fig. 1.3). This result indicates that the catalytic activity of the protease is necessary, if not sufficient, for normal processing explaining why proper processing occurrs in a foreign host. [It is interesting to note that α-lytic protease shows a strong preference to cleave peptides on the C-terminal side of small hydrophobic amino acid residues, with Ala being the best substrate (Bauer et al., 1981). That cleavage actually occurs *before* the Ala in the

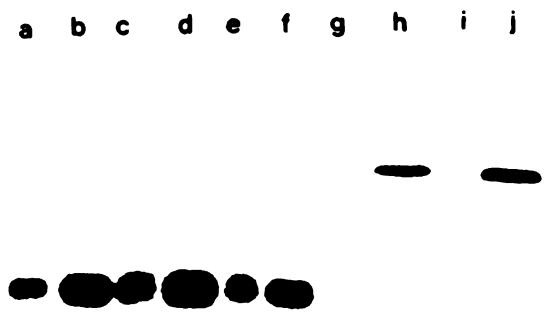


Fig.1. 3. Western blot of cell extracts of induced α- lytic protease constructs. Cultures of MH1 containing the various constructs were grown at 24-25°C unless otherwise indicated and induced for 3 days. All samples are the equivalent of 50 microliters of culture, except pALP5 grown at 37°C, which is equivalent to 100 microliters. a) 100ng of purified α-lytic protease from *L. enzymogenes*; b,c) pALP4 cell pellet, supernatant; d,e) pALP5 cell pellet, supernatant; f,g) pMALP2 cell pellet, supernatant; h,i) pALP5 grown at 37°C, cell pellet, supernatant; j) pALP5SA195 cell pellet.

Discussion

In this paper we have described the successful expression of α-lytic protease in *E. coli*, using the *pho*A promoter and signal sequence. Production of active protease, which accumulates in the extracellular medium, is dependent upon the presence of the large amino-terminal pro region that is part of the *L. enzymogenes* gene structure. Expression of the protease domain alone results in an inactive, but properly sized and apparently stable, protein. In fact, native gels comparing the migration of natural protease and the inactive protease fragment indicate that there are no gross structural anomalies associated with the absence of the pro region (data not shown). The fact that a mutation in the active site of α-lytic protease inhibits precursor processing indicates that α-lytic protease is responsible for its own processing. In all these respects, α- lytic protease mimics subtilisin E, the *Bacillus* serine protease which has an amino-terminal pro region of 77 amino acids (Stahl and Ferrari, 1984). Whether the relatively large size of the α-lytic precursor (199 amino acids) is indicative of some additional function in the native organism, such as transport across the second membrane of *L. enzymogenes*, is an interesting point of speculation. In this regard, it should be pointed out that when the protease domain alone is expressed, it remains cell associated, presumably in the periplasmic space, whereas expression of the preproenzyme leads to mature α-lytic protease accumulation in the medium. The general leakiness of the outer membrane seen in our expression of the preproenzyme argues against a specific transport mechanism being utilized in E. coli.

The phenomenon of temperature sensitivity in the production of active α-lytic protease in *E. coli* is particularly interesting when one considers the

stability of the mature product. In both *L. enzymogenes* and *E.coli* cultures, α-lytic protease is sufficiently stable to accumulate in the culture medium for several days, showing an unusual resistance to autolysis. In the case of the native enzyme, greater than half the catalytic activity is maintained at 66°C (Kaplan et al., 1970). Apparently, in the formation of the active stable product there is a step that is highly sensitive to changes in temperature.

Our current thinking is that the remarkable temperature stability and protease resistance of the active enzyme is generated at the expense of ease of folding. Both these aspects of stability may be explained by X-ray crystallographic analysis of the active enzyme. The enzyme structure shows an exceptional degree of structural rigidity (Fujinaga et al., 1985), suggesting that there is a high energy barrier between the folded and unfolded states. A high energy barrier could account for the inactive state of the protease domain when expressed alone, and for the apparent inability of the enzyme, once denatured, to refold (Dr. J. Richards, personal communication). By analogy with the concept of transition state stabilization in enzyme reactions, we hypothesize that the pro region would preferentially bind and stabilize the protease domain's "transition state" for folding, essentially acting as a template on which the mature enzyme finds its active conformation. Experiments to test this hypothesis are under way.

LIBRARY
Francisco
RARY
LIBRARY
Francisco
RARY
LIBRARY

FOOTNOTE

*For the purposes of describing mutations introduced into the mature α-lytic protease, we have adopted the chymotrypsin numbering scheme. For example, the Asp-His-Ser catalytic triad in α-lytic protease (amino acids 63,36,143) would be referenced by their homologous positions in chymotrypsin (amino acids 102, 57, 195). This allows a more direct comparison of our data with the exceedingly broad literature on protease structure and function.

Acknowledgements

We thank Dr. Roberta Parente for graciously agreeing to analyze the yeast extract for phosphate content. We also thank Dr. Marc Navre for helpful comments on the manuscript, and Dr. Susan Michaelis for many helpful discussions. This work has been supported by grants from the NSF Presidential Young Investigator Program, and the Howard Hughes Medical Institute. D.F. and A. F. were supported by NSF Graduate Fellowships.

Bachovchin, W. W., and J. D. Roberts. 1978. Nitrogen-15 nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of α-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases. J. Amer. Chem. Soc. **100**:8041-8097.

Bachovchin, W. W., R. Kaiser, J. H. Richards, and J. D. Roberts. 1981. Catalytic mechanism of serine proteases: reexamination of the pH dependence of the histidyl 1 $J_{3C2\text{-}H}$ coupling constant in the catalytic triad of α-lytic protease. Proc. Natl. Acad. Sci. USA **78**:7323-7326.

Bauer, C. A., G. D. Brayer, A. R. Sielecki, and M. N. G. James. 1981. The active site of α-lytic protease. Enzyme-substrate interactions. Eur. J. Biochemistry 120:289-294.

Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513-1523.

Bone, R., A. B. Shenvi, C. A. Kettner, and D. A. Agard. 1987. Serine protease mechanism: structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. Biochemistry **26**:7609-7614.

Brayer, G.D., L. T. J. Delbaere, and M. N. G. James. 1979. Molecular structure of the α-lytic protease from Myxobacter 495 at 2.8 Å resolution. J. Mol. Biol. **131**:743-775.

Brickman E. and J. Beckwith. 1975. Analysis of the regulation of

Escherichia coli alkaline phosphatase synthesis using deletions and ϕ80 transducing phages. J. Mol. Biol. **96**:307-316.

Burnette, W. H. 1981. Western blotting: Electrophoretic transfer of proteins from SDS-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. **112**:195-203.

Efstratiadis, A., J. N. Vournakis, H. Doris-Keller, G. Chaconas, D. K. Dougall, and F. C. Kafatos. 1977. End labeling of enzymatically decapped mRNA. Nucleic Acids Res. **4**:4165-4174.

Fujinaga, M., L.T.J. Delbaere, G. D. Brayer, and M. N. G. James. 1985. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. **183**:479-502.

Hall, M. H., L. Hereford, and I. Herskowitz. 1984. Targeting of E. coli beta-galactosidase to the nucleus in yeast. Cell **36**:1057-1065.

Hames, B. D., and D. Rickwood. 1981. Gel electrophoresis of proteins. IRL Press, Oxford, U.K.

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. Bact. **169**:3778-3784.

Hunkapiller, M. W., S. H. Smallcombe, D. R. Whitaker, and J. H. Richards. 1973. Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implications for the catalytic mechanism of serine proteases. Biochemistry **12**:4732-4743.

Ikemura, H. H. Takagi, and M. Inouye. 1987. Requirement of pro-

sequence for the production of active subtilisin E in *Escherichia coli*. J. Biol. Chem. **262**:7859-7864.

Inouye, H., S. Michaelis, A. Wright, and J. Beckwith. 1981. Cloning and restriction mapping of the alkaline phosphatase structural gene (*pho*A) of *Escherichia coli* and generation of deletion mutants in vitro. J. Bact. **146**:668-675.

Jacobs, M., M. Eliasson, M. Uhlen, and J. I. Flock. 1985. Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res. **13**:8914-8926.

Kaplan, H. V. B. Symonds, H. Dugas, and D. R. Whitaker. 1970. A comparison of properties of the α-lytic protease of Sorangium sp. and porcine elastase. Can. J. Biochem. 48:649-658.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.

Michaelis, S., L. Guarente, and J. Beckwith. 1983. In vitro construction and characterization of *pho*A-*lac*Z gene fusions in *Escherichia coli*. J. Bact. **154**:356-365.

Miller, J. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Morrison, D. A. 1979. Transformation and preservation of competent bacterial cells by freezing. Methods in Enzymol. **68**:326-331.

Naito, S., and Ueda, T. 1981. Affinity-purified anti-protein I antibody. J. Biol. Chem. **256**:10657-10663.

Neidhardt, F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. J. Bact. **119**:736-747.

Norrander, J., Kempe, T. and Messing, J. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed

LIBRARY
LIBRARY

mutagenesis. Gene **26**:101-106.

Norris, K., F. Norris, L. Christiansen, and N. Fiil. 1983. Efficient site-directed mutagenesis by simultaneous use of two primers. Nucleic Acids Res. 11:5103-5112.

Olson, M. O. J., N. Nagabhushan, M. Dzwiniel, L. B. Smillie, and D. R. Whitaker. 1970. Primary structure of α-lytic protease: a bacterial homologue of the pancreatic serine proteases. Nature **228**:438-442.

Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature **325**:458-462.

Robillard, G., and R. G. Shulman. 1974. High resolution nuclear magnetic resonance studies of the active site of chymotrypsin. I. The hydrogen bonded protons of the "charge relay" system. J. Mol. Biol. **86**:519-540.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-5467.

Shimada, H., M. Honjo, I. Mita, A. Nakayama, A. Akaoka, K. Manabe, and Y. Furutani. 1985. The nucleotide sequence and some properties of the neutral protease gene of *Bacillus amyloliquefaciens*. J. Biotechnology **2**:75-85.

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, in press.

Stahl, M. L., and E. Ferrari. 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived

LIBRARY
LIBRARY
LIBRARY

deletion mutation. J. Bacteriol. 158:411-418.

Struhl, K. 1985. A rapid method for creating recombinant DNA molecules. Biotechniques 3:452-453.

Takagi, M., T. Imanaka, and S. Aiba. 1985. Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus.* J. Bacteriol. **163**:824-831.

Vasantha, N., L. D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula. 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bact. 159:811-819.

Wells, J. A., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen. 1983. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis.* Nucleic Acids Res. **11**:7911-7925.

Whitaker, D. R. 1970. The α-lytic protease of a Myxobacterium. Methods in Enzymol. **19**:599-613.

Yang, M. Y., E. Ferrari, and D. J. Henner. 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. **160**:15-21.

Yanagida, N., T. Uozumi, and T. Beppu. 1986. Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli..* J. Bacteriol **166**:937-944.

From these data, it appears that a functional α-lytic protease active site is required for cleavage of the precursor into pro + protease regions. Before concluding that α-lytic is self processing, however, it is important to consider the possibility that the mutation destabilizes the native structure in such a way that the cleavage site is no longer accessible or recognized by the processing enzyme. Several facts argue against this possibility. Neither chemical modification of trypsin's active site serine to dehydroalanine nor mutagenesis of the active site serine appear to alter the structure of the rest of the protein. Mutation of the active site serine of subtilisin (in a system where its pro region is cleaved in trans by wild type molecules) also produces a correctly folded molecule (Fujishige et al., 1992). And finally, as discussed in Chapter 3, production of α-lytic protease SA195 in a complementation system (which precludes the processing step) results in an apparently folded molecule. Thus, without certainty but with a high probability of correctness, I conclude that α-lytic protease is self-processing.

This leaves two somewhat interrelated questions. The precursor, to our knowledge, is inactive. Nevertheless, it is self processing. This creates a paradox: how can an inactive molecule become active when its own activity is required for the activation process? Therefore, the first question that comes to mind is, how is the self-processing initiated? And second, is the processing event inter- or intramolecular? The initiation problem is circumvented if the event is intramolecular: each molecule then performs its own processing, possibly incapable of aaccepting any other substrate until this act has been performed. On the other hand, if the processing occurs intermolecularly, is there a particular protease that initiates processing? Or is there a particularly

protease sensitive site in the precursor, which, when cut, allows α-lytic to take over the processing itself? The native host, *Lysobacter enzymogenes*, produces many candidate enzymes for the initiation step. *E. coli* is less obliging, but it is possible that a periplasmic protease imperfectly cleaves the precursor to a partially active form, which in turn initiates the self processing cascade. In principle, the two mechanisms can be distinguished by their concentration dependence; the rate of intramolecular events is concentration independent, whereas the rate of intermolecular events varies with concentration depending on the number of molecules involved in the event. In our system, many factors conspire to make this approach complex: constraints imposed by allowable in vitro folding concentrations and the detection limits of our activity assay, as well as the fact that once cleaved, the pro region is a tight binding inhibitor of the protease. It may be feasible to address this question by performing refolding assays with wild type precursor dilutely bound to a column. By preventing intermolecular interactions, this protocol should only yield mature size, active molecules if the processing can occur intramolecularly.

Chapter 2: **Localization Studies**

The following chapter appeared in the Journal of Cell Biology, 1992, Volume 118, pages 33-42. Joy Silen created several of the expression vectors and Karen Smith performed all of the work concerning pro region deletions. The remainder of the work is mine. The localization study started as an attempt to facilitate purification of the precursor, and eventually revealed something interesting about protein targeting in Gram-negative cells.

Correct folding of α-lytic protease is required for its extracellular secretion from *E.coli*

Amy Fujishige, Karen Smith, Joy L. Silen[§], and David A. Agard*

Departments of Biochemistry & Biophysics, Pharmaceutical Chemistry and The Howard Hughes Medical Institute
University of California at San Francisco
San Francisco, CA 94143-0448

[§] Current address: Syva Co., 900 Arastradero Rd., Palo Alto, CA 94303

* To whom correspondence should be addressed

Running title:

Correlation of folding and extracellular secretion of α-lytic protease from *E. coli*

Abstract

α-Lytic protease is a bacterial serine protease of the trypsin family that is synthesized as a 39-kilodalton preproenzyme (Silen, J.L., McGrath, C.N., Smith, K.R. and D.A. Agard. 1988. *Gene* 69:237-244). The 198 amino acid mature protease is secreted into the culture medium by the native host, *Lysobacter enzymogenes* (Whitaker, D.R. 1970. *Meth. Enz.* 19:599-613). Expression experiments in *E. coli* revealed that the 166 amino acid pro region is transiently required either *in cis* (Silen, J.L., Frank, D., Fujishige, A., and D.A. Agard. 1989. *J. Bact.* 171:1320-1325) or *in trans* (Silen, J.L. and D.A. Agard. 1989. *Nature* 341:462-464) for the proper folding and extracellular accumulation of the enzyme. The maturation process is temperature sensitive in *E. coli*; unprocessed precursor accumulates in the cells at temperatures above 30°C (Silen, J.L., McGrath, C.N., Smith, K.R. and D.A. Agard. 1988. *Gene* 69:237-244). Here we show that full-length precursor produced at nonpermissive temperatures is tightly associated with the *E. coli* outer membrane. The active site mutant Ser 195->Ala (SA195), which is incapable of self-processing, also accumulates as a precursor in the outer membrane, even when expressed at permissive temperatures. When the protease domain is expressed in the absence of the pro region, the misfolded, inactive protease also cofractionates with the outer membrane. However, when the folding requirement for either wild type or mutant protease domains is provided by expressing the pro region *in trans*, both are efficiently secreted into the extracellular medium. Attempts to

separate folding and secretion functions by extensive deletion mutagenesis within the pro region were unsuccessful. Taken together, these results suggest that only properly folded and processed forms of α-lytic protease are efficiently transported to the medium.

Introduction

Protein targeting has been a topic of intense study in recent years (Pugsley, 1989). In Gram-negative bacteria, there are four noncytoplasmic destinations: the inner membrane, the periplasm, the outer membrane, and the extracellular medium. While targeting to the inner membrane or the periplasmic space is fairly well understood, general mechanisms and cues for targeting proteins to the outer membrane or the extracellular medium have not yet been identified (for review see Model and Russel, 1990). It is unclear to what extent transport to the extracellular medium and to the outer membrane share mechanistic features with transport to the inner membrane and periplasm. In certain cases, extracellular proteins seem to translocate without a signal sequence, and it has been suggested that they do so directly via the Bayer junctions (Filloux et al., 1990; Mackman et al., 1985; Wandersman and Delepelaire, 1990). In other cases, proteins translocate first to the periplasm using the signal sequence-dependent pathway and are subsequently translocated to the outer membrane (Sen and Nikaido, 1990) or external medium with the help of accessory proteins. The recent demonstration that fourteen genes of the *pul* operon are required for the extracellular secretion of the *Klebsiella pneumoniae* enzyme pullulanase (Pugsley et al. 1990) highlights the potential complexity of extracellular transport systems. However, the finding that two of the *xcp* genes required for protein secretion from *Pseudomonas aeruginosa* show significant homology to *pul* genes (Filloux et al., 1990) suggests that at least some aspects of extracellular transport from Gram-negative bacteria may share common mechanisms.

Many extracellularly secreted proteases bear large amino- or carboxy-terminal pro regions. Where examined, secretion of these molecules to the

extracellular medium appears to be dependent on the presence of a signal sequence as well as the presence of the pro region. The *Serratia marcescens* serine protease is synthesized as a 112-kilodalton preproenzyme, whose amino-terminal signal sequence and 52-kilodalton carboxy-terminal pro region are cleaved during export through the inner and outer membranes, respectively (Miyazaki et al., 1989). Similarly, aqualysin I is produced by *Thermophilus aquaticus* with an amino-terminal signal sequence and pro region in addition to a large carboxy-terminal pro region (Terada et al., 1990). Both of these proteases can be expressed in *E. coli*, and become processed and secreted under proper conditions. For the IgA protease of *Neisseria gonorrhoeae*, it has been proposed that the pro region forms a pore in the outer membrane through which the protease is translocated. Following translocation, the mature enzyme is released from the pro region by proteolysis (Pohlner et al., 1987).

α-Lytic protease is an extracellular serine protease from the Gram-negative soil bacterium *Lysobacter enzymogenes*. The protein has been extensively studied as a model system for investigating serine protease mechanism (Hunkapiller et al., 1973; Bachovchin et al., 1988; Bone et al., 1987) and for understanding the structural basis for enzyme specificity (Bone et al., 1989a; Bone et al., 1989b, Bone et al., 1991). α–Lytic protease is synthesized as a preproenzyme (Silen et al., 1988); expression experiments in *E. coli* have revealed that the 166 amino acid pro region plays an obligatory role in the folding of the 198 amino acid protease domain (Silen et al., 1989). For this reason, the mechanism of folding of α-lytic protease has been examined. Constructs lacking the pro region (Δ-pro-α-lytic protease) give rise to inactive α-lytic protease (Silen et al., 1989). Remarkably, it is possible to complement the folding defect of this Δ-pro molecule by *in vivo* co-expression of the pro region *in trans* (Silen and Agard, 1989). Similarly, *in vitro* experiments

demonstrate that it is possible to refold denatured α-lytic protease only in the presence of the pro region (Baker et al., 1992). As in the *in vivo* system, the pro region can effect folding in either the presence or the absence of a covalent attachment to the protease region. Moreover, the *in vitro* folding is temperature-sensitive; folding is greatly retarded at temperatures above 30°C (Fujishige, A. , D. Baker and D.A. Agard, manuscript in preparation).

When wild type α-lytic protease is expressed in *E. coli*, the mature (proteolytically processed and active) protease first appears in the periplasm and then accumulates in the medium at temperatures below 30°C. When the same construct is expressed at temperatures above 30°C, where it is incompetent to fold, an inactive precursor accumulates in the cells. Inactivation of the protease by mutation of the catalytic serine residue results in a similar accumulation of cell-associated precursor even at permissive temperatures, indicating that α-lytic protease is self-processing in *E.coli*. The misfolded Δ-pro-α-lytic protease is also found to be cell-associated, independent of expression temperature (Silen et al., 1989).

In this paper, we provide evidence that the Δ-pro-α-lytic protease, mutant inactive precursor, and the wild type precursor synthesized at the restrictive temperature are all tightly associated with the *E. coli* outer membrane. In addition, deletion analysis of the pro region (performed by Karen Smith) reveals that the folding and the secretion functions are not easily separable and that only proteolytically active molecules are efficiently secreted across the outer membrane. A folded but inactive protease produced by mutation of the protease active site in complementation with a wild type pro region is efficiently transported to the medium, indicating that enzyme activity *per se* is not required for secretion. We therefore propose that, unlike translocation through the cytoplasmic or the mitochondrial membrane, which

requires that the nascent protein be in an unfolded state (Randall and Hardy, 1986; Eilers and Schatz 1986), efficient translocation through the outer membrane appears to require that the protein be correctly folded.

Materials and Methods

Plasmid constructions-- pALP5, pALP5-SA195, and pMALP2 are pBR329 derivatives that have been described previously (Silen et al., 1989). pALP5 contains a copy of the α-lytic proenzyme under the control of the *phoA* promoter and fused to the *phoA* signal sequence. pALP5-SA195 is the same vector with the active site serine 195 (chymotrypsin numbering, James et al., 1978) mutated to alanine to inactivate the enzyme. pMALP2 contains the protease region fused directly to the *phoA* signal sequence, producing the inactive protease referred to as Δ-pro-α-lytic protease.

pALP6 was created to introduce unique restriction sites in pALP5 in order to conveniently subclone and express deletion mutants of the pro region (see below). It was constructed by ligating the filled-in *Eag* I/*Xba* I restriction fragment of pALP5 containing the α-lytic protease gene into the *Sma* I site of the *phoA* expression vector pDBR2 (Silen et al., 1988) that had previously been digested with *Not* I and *Nar* I, filled in with Klenow (NEB) and religated. pALP7junk was generated to allow deletion mutagenesis at the N-terminus of the pro-region without subcloning. The 'junk' portion protects the promoter and is removed after *Bal* 31 deletion and prior to ligations (see below). It was produced in two cloning steps. First the filled-in *Eag* I/*Xba* I restriction fragment of pALP5 was cloned into the filled-in *Xma* I site of pDBR2 (*Not* I/*Nar* I dropout) creating the out-of-frame construct pALP7. Then the 686bp filled-in *Bgl* II/*Acc* I fragment of m13mp18 was cloned into the pALP7 filled-in *Xma* I site in the minus orientation.

For the construction of the pro-region vector pPRO2, the *Eco*47III/*Bal* I fragment of pPRO1 (Silen and Agard, 1989) was ligated into pDBR2 from which the *Sty* I sites had been removed by religation after treatment with *Nco* I, *Sty* I, and Klenow fragment. pMALP4 is identical to pMALP2 (Silen et al., 1989)

except that unique *Mlu* I and *Xho* I sites have been introduced by replacing the *Bal* I/*Nco* I fragment with the analogous fragment from pALP5 (Silen et al., 1989). For the construction of the complementation vector pCOMP4, the coding region for the protease was excised from pMALP4 by cutting first with *Nco* I, filling in with Klenow fragment, and then cutting with *Sph* I. This fragment was ligated into pPRO2 that had been treated with *Not* I, Klenow, and *Sph* I. For the construction of pCOMP4-SA195, the *Mlu* I/*Xho* I fragment of pALP5-SA195 (Silen et al., 1989) was ligated into the analogous sites of pCOMP4 (Figure 2.1).

Deletion Mutagenesis of the Pro Region-- Silent restriction sites within the coding sequence for the pro region were made by oligonucleotide-directed mutagenesis as described by Kunkel (1985) using an m13mp18 construct containing the pALP5 *BamH* I fragment in the *BamH* I site of m13mp18. The silent restriction sites and their respective mutagenic oligonucleotides are: *Bgl* I,5'-GATGCCCAGATCTCGCTGCATTGCG -3'; *Hind* III,TGGGTGCGCGCAAGCTTTTCGGTCT-3'; *Hind* III#2, AGCTTGAAGCTTCCGTCTTCGTTGC; and *Dra* III, TTCCACCCCACCCAGTG-TGCTCGAC. The oligonucleotides were synthesized at Operon Tech., Inc. For internal sites in the pro region, the mutated *BamH* I fragments were ligated individually into pALP6.

Bal 31 exonuclease digestions were carried out as follows. 10μ g plasmid was digested to completion with the endonuclease for the targeted site in a 50μl reaction. The reaction was adjusted to 90ml containing 12.5mM $MgCl_2$, 12.5mM $CaCl_2$, 50mM NaCl, 20mM Tris (pH8), and 1mM EDTA[1]; divided into three samples, each with a different level (0.01u, 0.1u, and 1.0u) of *Bal* 31 Exonuclease (BMB). Each sample was incubated at 30°C and 7.5μ l aliquots were withdrawn after 1, 2, 5, and 10 minutes. Reactions were stopped by addition of 10μl 0.1M EDTA/0.1M [ethylenebis(oxyethylnitrilo)]tetraacetic acid (EGTA). Time points were phenol/chloroform extracted, ethanol precipitated, and

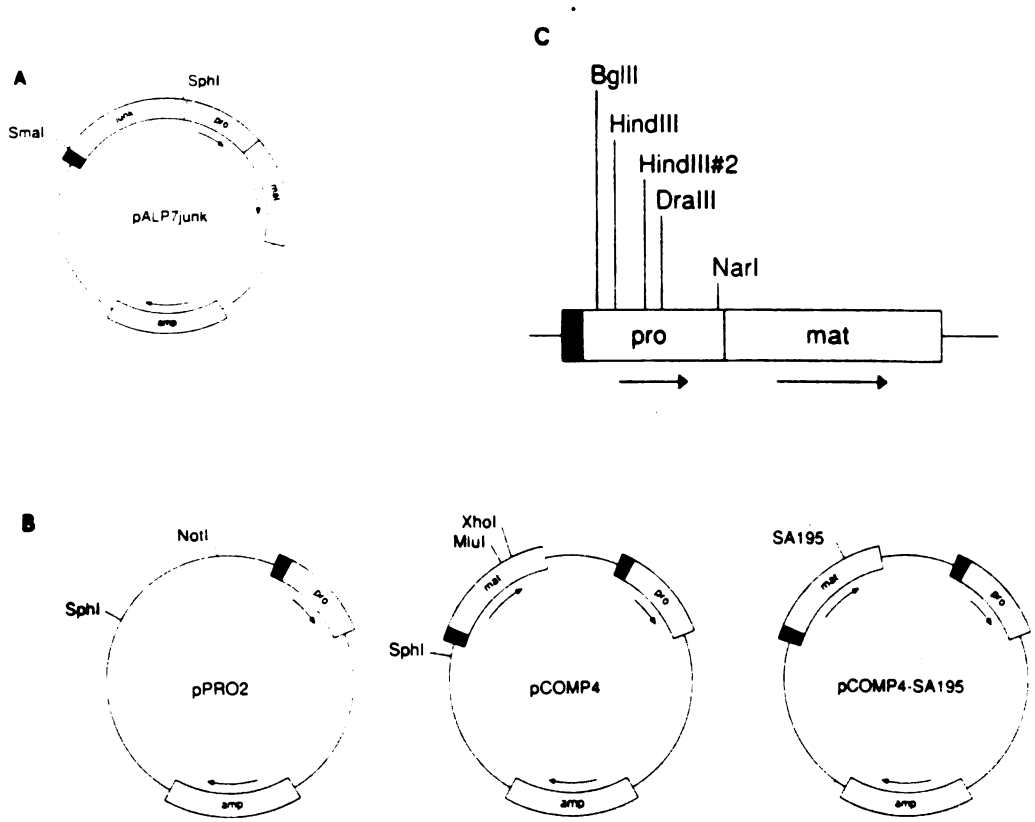


Figure 2.1. **Deletion and Expression vectors** (see Materials and Methods for construction). The shaded areas represent the *PhoA* promoter and signal sequence. The sections designated 'pro' refer to the pro region open reading frame, and those marked 'mat' refer to the sequence that encodes mature size protease. *A,* pALP7junk vector for N-terminal deletions of the pro region. The section designated 'junk' protects the promoter and signal sequence during bidirectional *Bal* 31 digestion starting at the *Sph I* site. It is subsequently removed by digestion with *Sma I,* and the vector is cleanly reclosed for transformation and expression. *B,* Expression vectors pPRO2,pCOMP4 and pCOMP4-SA195. pCOMP4 and pCOMP4-SA195 are identical except that the sequence coding for the active site serine has been mutated to encode alanine in pCOMP4-SA195. *C,* Target sites for pro region deletions. Schematic diagram indicating the positions of silent restriction sites within the pro region. The sites were introduced individually into the expression vector pALP6.

treated with Klenow. To determine the extent of *Bal* 31 digestion, a portion of each sample was analyzed by gel electrophoresis; appropriate samples were ligated and transformed into *E. coli* MH1 cells. For N-terminal deletions of the pro region, *Bal* 31 digestions were performed at the *Sph* I site of pALP7junk. After the extraction and precipitation, the DNA was digested with a second endonuclease, *Sma* I, treated with Klenow, ligated, and transformed into *E. coli* MH1 cells .

The deletion mutants were prescreened for expression by colony immunoblot. MH1 cells transformed with mutant-bearing plasmids were plated on LB plates. After 12 hours at 37°C, a nitrocellulose circle (Schleicher & Schuell, 0.45μm pore) was placed on top of the growing colonies. The colonies were allowed to adsorb to the filter for one minute before carefully removing the nitrocellulose filter. The filter was then placed colony side up on an inducing plate (1.5% agar containing the modified MOPS medium described below). The colonies were allowed to grow and express protein for twelve to thirty six hours at 25°C. At that point, the cells were washed off the filters with blot buffer and subjected to immunogold and silver staining with Janssen AuroProbe BLplus and IntenSE BL reagents using affinity-purified rabbit anti-α- lytic protease antibody (Silen et al., 1989) as the primary antibody. Colonies producing protease were then subjected to sequencing, activity analysis, and gel electrophoresis. Activity assays were performed as described (Silen et al., 1989) after treatment with trypsin to remove bound pro region (see below).

Double Stranded Sequencing --DNA was extracted from 5 ml overnight culture by the alkaline extraction method of Birnboim and Doly (1979). The DNA was further purified by polyethylene glycol precipitation as described by Hattori and Sakaki (1986). Double stranded sequencing was performed on the

entire DNA preparation according to the methods outlined in Tabor and Richardson (1987).

Strains, media and expression-- *E.coli* MH1 (*araD139 ΔlacX74 galU galK hsr rpsL;* 13) was used as the host for expression of all forms of α-lytic protease described above. LB medium containing 100ug/ml carbenecillin was used to maintain expression of plasmids. For *phoA*-directed expression, single colonies were picked into LB medium containing 100ug/ml carbenecillin and allowed to grow overnight at 37°C. The cells were then pelleted, washed three times in morpholinepropanesulphonic acid (MOPS) medium (Neidhardt et al., 1974), and diluted 1:50 into MOPS medium supplemented with 0.2% glucose, 0.15% vitamin free Casamino Acids (Difco Laboratories, Detroit, MI), and 0.05% Yeast Extract (Difco). This modified MOPS medium provides a low phosphate environment that allows the cells to reach high density before induction by phosphate depletion. The cells were grown at either 37°C or 22°C as indicated, and harvested after 36 hours.

Initial Localization and Solubility Studies-- Whole cells were examined at 1000x magnification using wet mount phase contrast microscopy. Although all cells carrying α-lytic-bearing plasmids were irregular in form (elongated, some incomplete septation), those that produce cell-associated forms did not differ from their secreting counterparts. Large refractile bodies or inclusions were not observed.

Cells and media were separated by centrifugation. The presence of α-lytic protease in either cells or supernatants was determined by immunoblots of sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE, see below). Cells were resuspended in Laemmli sample buffer (Laemmli, 1970) and boiled for five minutes to effect lysis. Supernatants were dialysed and concentrated where necessary, and boiled in sample buffer for one minute.

The periplasm was extracted by osmotic shock using the method of Neu and Heppel (1965). Alternatively, complete cell lysis was effected by treatment in 8% sucrose, 45mM EDTA, 5% Triton X-100, 50mM Tris, pH 8.0 with 10 mg/ml lysozyme. The lysed cells were further extracted with 2.5% β-octyl glucoside or 0.2% deoxycholate in 5mM potassium phosphate, pH 7. In separate studies, whole or lysed cells were treated with 1%Triton X-100 in the absence or presence of 5mM EDTA (differential extraction of inner and outer membranes). In other analyses, the lysed cells were extracted with buffers from pH 4 to pH 10.5. After each extraction method mentioned, the sample was centrifuged. The supernatants were dialysed where necessary, and both the supernatant and the pellets were analysed by SDS-PAGE followed by immunoblotting (see below).

Membrane preparation and fractionation--A crude membrane preparation was made by the method of Mizushima and Yamada (1975). The membranes were resuspended in 1% EDTA at pH 7, or in 1% EDTA at pH7 with 200mM NaCl, or in 1% EDTA at pH7 with 1M NaCl, or in 1% EDTA with 200mM Na_2CO_3 at pH 10.5. Sucrose gradients were poured in 5% steps, 35-55% (w/w) sucrose. For each sample, the sucrose gradient contained 5mM EDTA and the same salt and pH as the sample resuspension buffer. 100-400 μl treated membranes were separated by isopycnic sucrose density gradient centrifugation at 125,000xg for 4 hours using a Beckman SW60 rotor. 200 μl fractions were collected from the top of each gradient. The bottom of each tube was washed with Laemmli sample buffer and is referred to as the gradient pellet. Sucrose concentration of the fractions was assessed by refractive index. Similar studies were performed with 1% Triton X-100 or 1M urea in the resuspension and sucrose buffers.

For flotation experiments, the crude membranes were mixed with sucrose and the appropriate buffer to produce 55% sucrose samples. These were placed on the bottom of the SW60 tube and the step gradients were carefully overlaid on top. Each flotation experiment was performed with an analogous density gradient experiment. The pairs were centrifuged at 125,000xg for 5 hours, at which time the membrane bands appeared to be in the same location within the experimental error and variations due to applied sample density and volume.

Assay for outer membranes-- 50 μl of each fraction was diluted tenfold to 500 μl with distilled water and the membranes were recovered by centrifugation at 35,000xg for 45 minutes. The pellets were assayed for the lipopolysaccharide marker, 2-keto-3-deoxyoctonate (KDO), by the microassay of Karkhanis et al.(1978).

Trypsin accessibility-- precursor produced by expression of the wild type construct (pALP5) at 37°C was tested for trypsin sensitivity in whole cells versus EDTA-permeabilized cells. For EDTA-permeabilization, cells were washed twice with ice cold 10mM Tris, pH 8.0. Next, they were resuspended in cold 5mM EDTA, 10mM Tris pH 8 and placed on ice for 15 minutes. The treated cells were microfuged and carefully resuspended in 0.5 mM EDTA, 10 mM Tris, pH 8.0 using the original culture volume. EDTA-permeabilized and whole cells (in culture medium) were treated with 10 μ g/ml trypsin on ice for 15 minutes. The cells were recovered by centrifugation and immediately resuspended in preheated Laemmli sample buffer that had been adjusted to pH 4.5 in order to stop the reaction. The samples were boiled for 5 minutes and microfuged for one minute before analysis by immunoblots of SDS gels.

Gel Electrophoresis of Proteins-- SDS-PAGE was carried out on a Hoefer minigel apparatus. Samples were combined with Laemmli sample buffer,

boiled for one minute, and microfuged for one minute. The supernatant was run on a discontinuous (4%/12%) SDS-polyacrylamide gel. Native-PAGE was carried out at low pH on discontinuous, nondenaturing gels prepared and run as described by Hames and Rickwood (1981). pCOMP4-SA195 samples were trypsinized briefly to remove bound pro region (see below).

Protein Quantitation and Immunoblot Analysis – The contents of SDS polyacrylamide gels were transferred to nitrocellulose using an ABN polyblot apparatus. For native gels, the same apparatus was employed except that the low pH gel electrophoresis running buffer was used for electrotransfer with reverse polarity from that used with SDS gels. Immunoblots were probed with affinity-purified rabbit anti-α-lytic protease antibody or an antibody directed against a fusion protein of the pro region with glutathione transferase (Baker et al., 1991) followed by [^{125}I] protein A (>300μCi/μg; ICN Pharmaceuticals, Inc., Irvine, CA) or alkaline phosphatase conjugated-goat-anti-rabbit antibody (Bio-Rad Laboratories, Richmond, CA). For relative quantitation of α-lytic protease in the sucrose gradients, the lanes of [^{125}I] protein A probed immunoblots were counted dry on a Beckman Gamma 8000 scintillation counter.

Results

Localization Studies--My initial attempts to purify the various cell-associated forms of α-lytic protease (Figure 2.2) revealed that, despite the presence of a signal sequence, the molecules were not present in periplasmic extracts. Instead, they were found in the pellet rather than the supernatant of total cell lysates. Although many extraction methods were tried, only those that disrupt strong protein-protein interactions (8M urea, 6M guanidine, SDS) were effective in releasing α-lytic protease from the cell pellet. Nonionic detergents such as β-octyl glucoside and Triton X-100, which are effective in extracting integral membrane proteins of the inner membrane but not usually those of the outer membrane, did not extract the cell associated forms of α-lytic protease. Deoxycholate solubilized only a small fraction of cell associated precursor. Variation of pH (5-10.5) was similarly ineffective in solubilizing the cell-associated forms. Inclusion bodies were not detectable by phase contrast microscopy (data not shown). These results led us to believe that the molecules might be associated with the outer membrane.

Examination of a partially-purified crude membrane fraction (as defined by Mizushima and Yamada, 1975) confirmed that the wild type precursor to α-lytic protease (produced at 37°C) was present almost exclusively in the membrane fraction. Notably, the wild type precursor accumulates in *E. coli* only at temperatures above the natural growth temperature of the native host, *Lysobacter enzymogenes*. In order to determine whether the precursor location was solely due to growth temperature, we examined the localization of an active site mutant, pALP5-SA195, which had previously been shown to accumulate as a cell associated, high molecular weight precursor even at temperatures that permit processing and secretion of the wild type protease

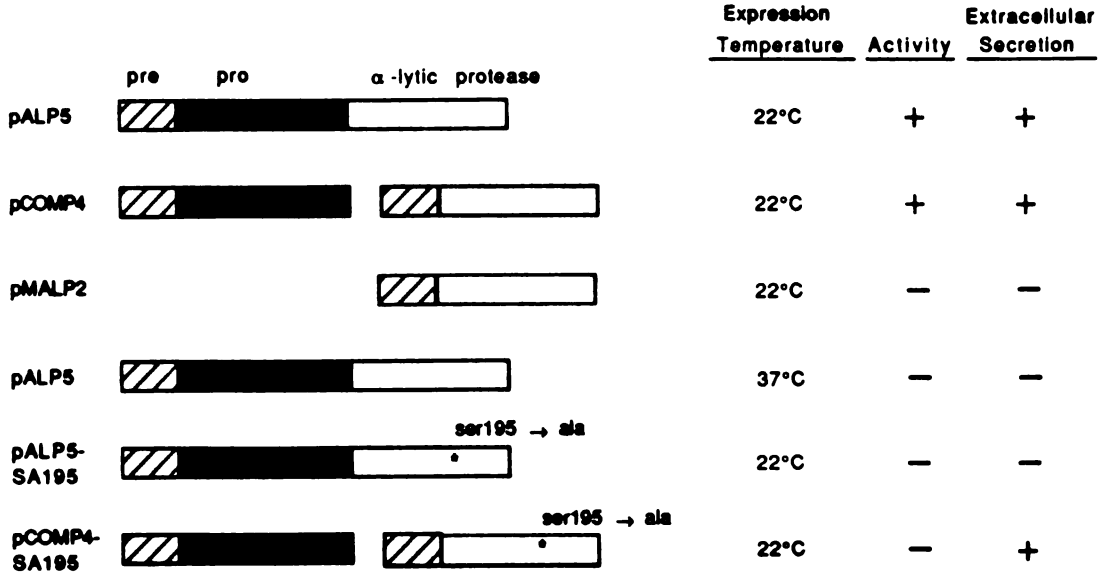


Figure 2.2. **Characteristics of expression constructs.** Cells and media were subjected to α-lytic protease assays as well as immunoblot analyses. The wild type construct (pALP5) expressed at permissive temperatures allows proper folding of the protease and cleavage of the pro region. α-lytic protease activity is required for the cleavage. Both the pro region and the mature protease are secreted into the medium. Physical linkage of the pro region is not required for proper folding and secretion, as shown by the complementation construct, pCOMP4. Deletion of the pro region (pMALP2) results in an inactive, cell associated molecule. Uncleaved precursors accumulate in the cell upon expression of wild type at nonpermissive temperatures or by mutation of the active site serine (pALP5-SA195) with growth at permissive temperatures. Mutation of the active site serine in the complementation construct (pCOMP4-SA195) allows secretion of the mature, folded, but inactive protease region.

(Silen et al., 1989). The mutant precursor was also found exclusively in the membrane fraction. Since both of these molecules retain a covalently attached pro region, we considered the possibility that the pro region mediated all interactions with the membrane. Therefore, we examined the location of the the wild type protease expressed with the *phoA* signal sequence but without the pro region (Δ-pro-α-lytic protease, pMALP2), a construct originally created by Joy Silen (a staff research assistant in the lab) and later modified by a rotation student, Dan Frank. This molecule is also inactive and cell associated, but appears by SDS-PAGE to have its signal sequence removed (Silen et al., 1989). Despite the absence of the large pro region, this molecule was also found in the crude membrane fraction.

Separation of the crude membrane fractions into inner and outer membranes on sucrose density gradients revealed that the precursor produced at 37°C, the active site mutant produced at 22°C, and the Δ-pro α-lytic protease produced at 22°C cofractionated with the outer membrane. Two major membrane bands were visible in the gradients: the lighter band, corresponding to inner membranes were observed at 45 to 48 % sucrose, while the denser band consistently migrated between 52 and 54 % sucrose. In one case, a ghost band was observed at 39% sucrose. KDO (2-keto-3-deoxyoctonate) assays verified that the densest membrane band contained outer membrane lipopolysaccharide, whereas the lighter bands did not. Immunoblots revealed that each of the cell associated forms was present only in those fractions that contained outer membrane material (Figures 2.3, 2.4, 2.5). The same result was obtained when the crude membrane fraction was loaded on the bottom of the gradient (data not shown) and allowed to float upwards, indicating that trapping was not causing the precursor to migrate anomalously.

A

7 8 9 10 11 12 13 14 15 16 17 P

B

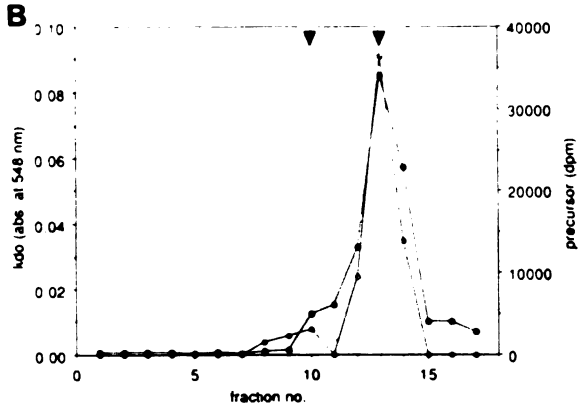


Figure 2.3. **Location of the proenzyme produced at 37°C.** Prepro-α-lytic protease was expressed at 37°C from plasmid pALP5 in the *E.coli* host cell, MH1. The inner and outer membranes were separated by isopycnic sucrose density gradient centrifugation. Fractions were collected from the top of the gradient. Major membrane bands were visible at 39, 46 (arrow) and 53% (arrrow) sucrose. *A*, autoradiograms of the fractions probed for α-lytic protease. The proteins of each fraction were separated by SDS-PAGE, transferred to nitrocellulose and probed first with anti-α-lytic protease antibody and then with [^{125}I] Protein A. All of the precursor containing fractions are shown; no signal was observed elsewhere. The emptied gradient tube was washed in Laemmli sample buffer to solubilize any proteins that might have pelleted at the bottom of the tube; this fraction is marked "P." *B*, the outer membrane content of each fraction as determined by KDO assay (open circles) , and relative quantification of the α-lytic precursor content as determined by gamma count of individual lanes of the immunoblot shown above (closed circles).

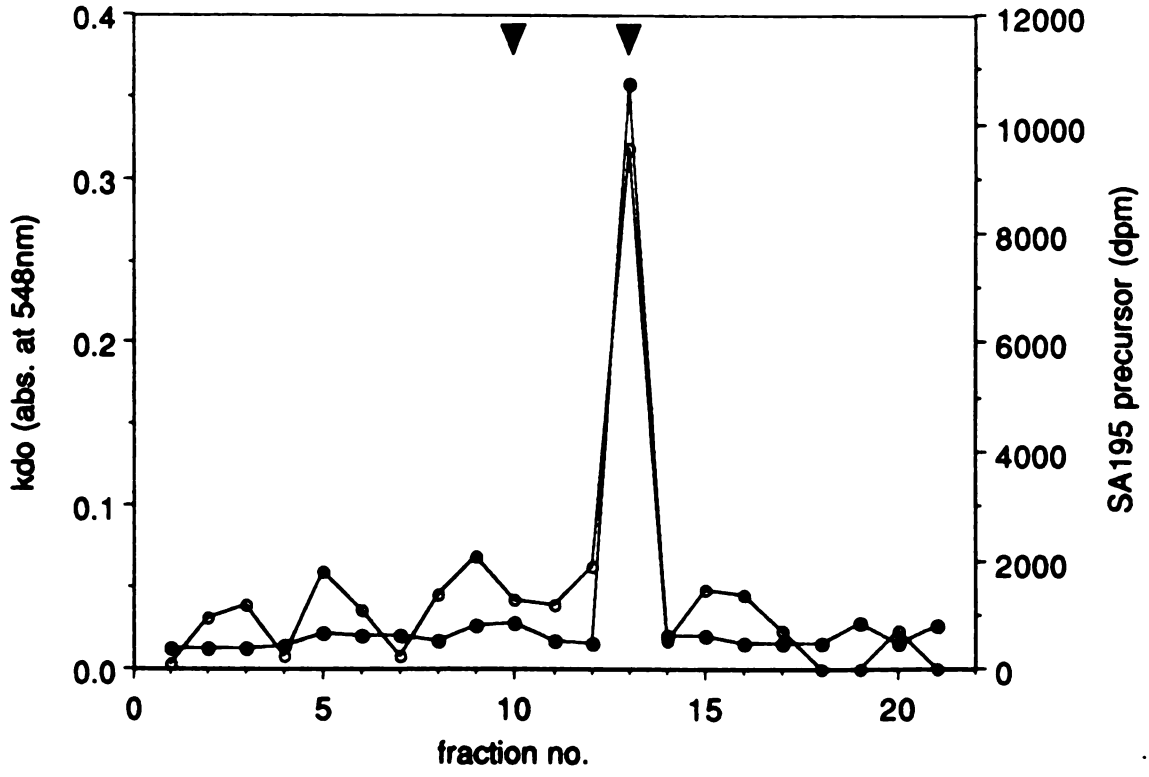


Figure 2.4. **Location of the active site mutant precursor produced at 22°C.** The active site mutant produced from vector pALP5-SA195 gives rise to a precursor that cofractionates with *E. coli* outer membranes. Crude membranes were fractionated on sucrose density gradients containing 5 mM EDTA, pH 7.0. Fractions were taken from the top of each gradient and assayed for KDO content (open circles) or α-lytic proenzyme content (closed circles) as above. Major membrane bands were visible at 46 and 52% sucrose, as indicated by arrows. Similar results were obtained using high salt or high pH (data not shown).

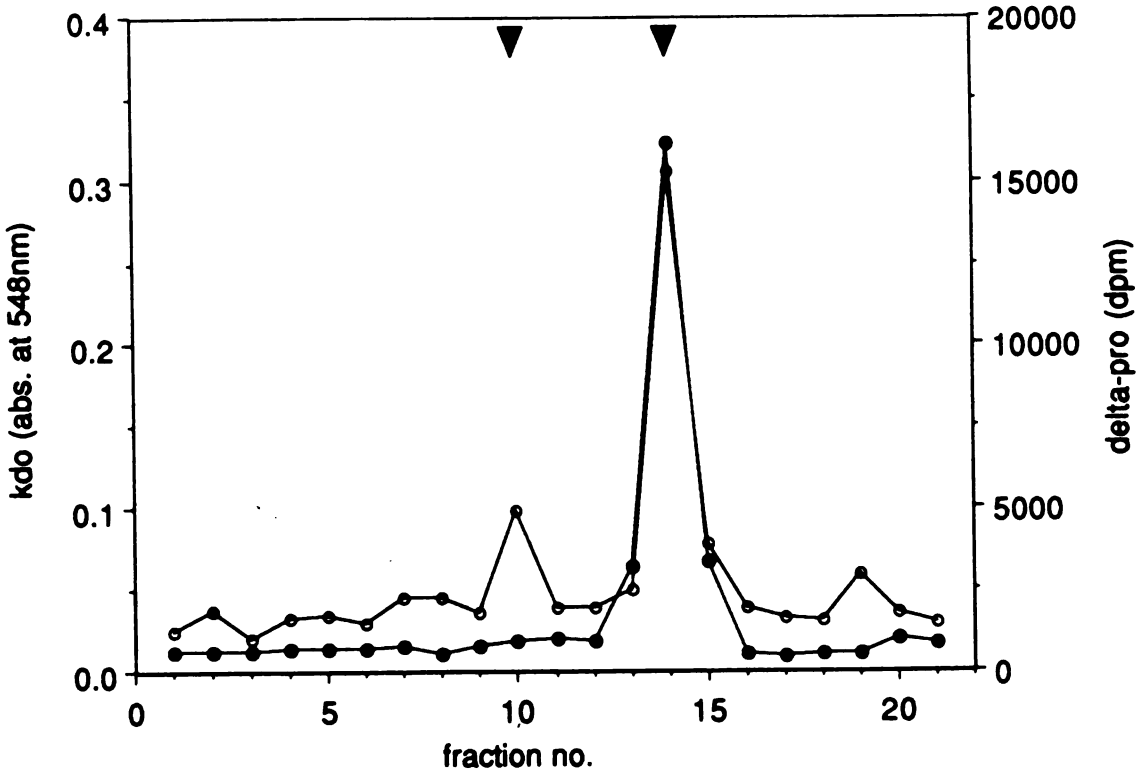


Figure 2.5. **Location of Δ-pro-α-lytic protease.** pMALP2 has a signal sequence fused directly to the protease region, and lacks the pro region altogether. The protein produced from this vector is also found in the outer membrane of *E.coli.* Membranes were fractionated in sucrose density gradients containing 5mM EDTA, pH 7.0. Fractions were taken from the top of each gradient and assayed for KDO content (open circles) or α-lytic protease content (closed circles) as above. Major membrane bands were visible at 48 and 52% sucrose, as indicated by arrows.

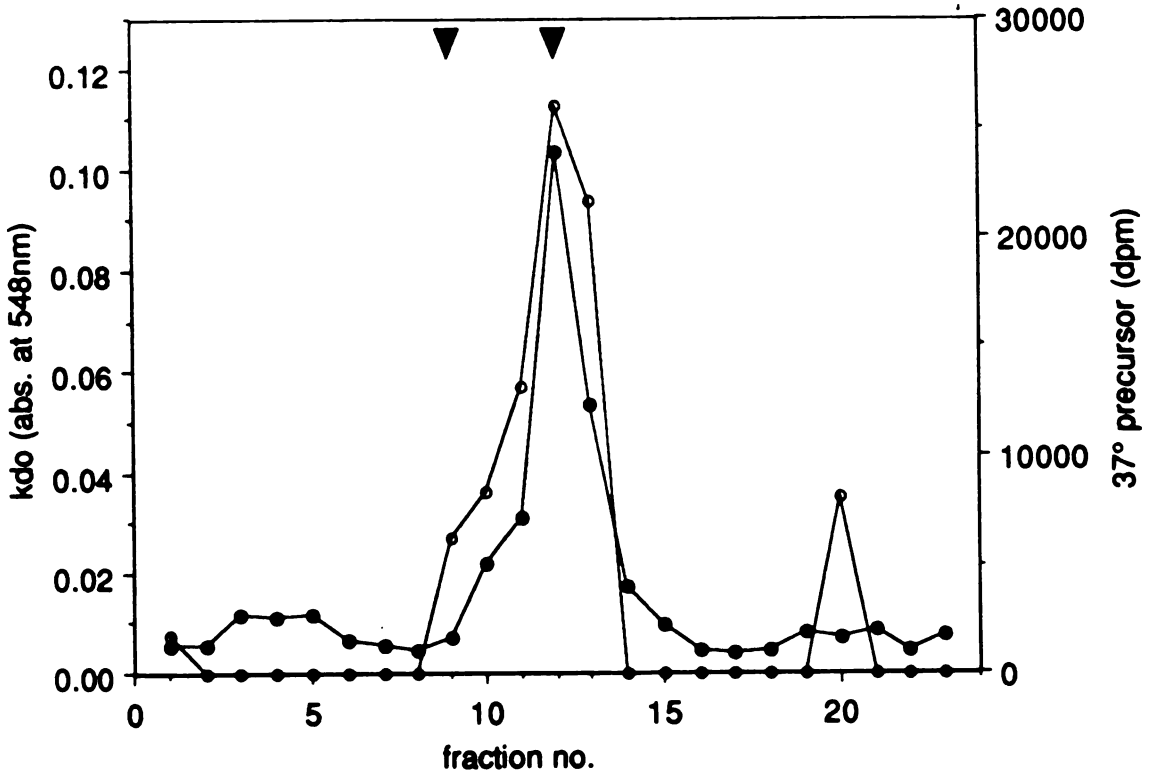


Figure 2.6. **Fractionation at high pH.** Prepro-α-lytic protease was expressed at 37°C and prepared as above. Crude membranes were resuspended in 200 mM Na_2CO_3, 5 mM EDTA, pH 10.5 and loaded on sucrose gradients made up in the same buffer. KDO content (open circles); α-lytic proenzyme content (closed circles). Major membrane bands were visible at 46 and 54% sucrose, as indicated by arrows. Similar results were obtained for the active site mutant (palp5-SA195) and Δ-pro-α-lytic protease (pMALP2) expressed at 22°C (data not shown).

In addition, we attempted to dislodge the Δ-pro-α-lytic protease from the membrane by mild treatments that are generally nondenaturing to proteins but that could disrupt weak associations with the membrane or membrane components. We therefore included 1% Triton X-100 or 1M urea in the resuspension of the crude membranes as well as in the sucrose gradient itself. Neither treatment altered the comigration of Δ-pro-α-lytic protease with the outer membrane (data not shown).

Significantly, neither the precursors nor the Δ-pro α-lytic protease could be dislodged from the outer membrane by treatment with either high pH (200mM Na_2CO_3, pH 10.5, Figure 2.6), or by salt (200 mM or 1M NaCl, data not shown). These procedures are known to be effective in the extraction of peripheral membrane proteins. we conclude that the various cell associated forms, which are strongly cationic, are not associated with the outer membrane solely by ionic interactions with negatively charged outer membrane components, and that their association with the outer membrane is strong. Treatment with Triton X-100 in conjunction with EDTA appeared to deplete the outer membrane of all three cellular forms of α-lytic protease, but did not effect complete extraction (data not shown). This result is consistent with the behaviour of outer membrane proteins OmpA, OmpC, and OmpF, which are partially extracted (~50%) by Triton-EDTA (Schnaitman, C.A. 1974; Hindennach and Henning, 1975) and outer membrane-localized fusions of LamB-LacZ, which are extracted to varying degrees by this treatment (Hall et al., 1982).

The accumulation of the mutant form of α-lytic protease precursor (pALP5-SA195) in the outer membrane of *E.coli* suggests that the association is not a simple artifact of growth temperature, but is likely to reflect a physical property of the precursor. Furthermore, strong association with the

outer membrane can occur even in the absence of the pro region. This indicates that the targeting signals are located in the protease region, or that the targeting signals are redundantly present in both the pro and the protease regions.

The precursor produced at 37°C is degraded by endogenous proteases but was only mildly accessible to trypsin that was added to intact cells (Fig. 2.7). By contrast, noticeable loss of the high molecular weight precursor was observed after fifteen minutes of trypsinization when the outer membrane of the cells was first permeabilized by treatment with EDTA (Fig. 2.8). A fragment of intermediate size between precursor and mature forms of α-lytic protease appeared and persisted. This suggests that a region of the precursor is protected from proteolysis. Such protection could occur from folding, interaction with the membrane, or tight interaction with another protein.

Correlation of Secretion and Folding -- The pro region has apparent roles in both folding and secretion (Silen et al., 1989). We sought to distinguish these two functions by deletion analysis. In work performed by Karen Smith (a staff research assistant in the lab), six areas throughout the pro region were targeted for deletion mutagenesis in an attempt to find a minimal unit responsible for either folding or secretion or to disable one, but not both functions. Silent restriction sites were introduced within the pro region; changes in α-lytic protease expression levels due to the resulting codon changes were not observed. Deletions at the restriction sites did, however, affect accumulation of mature protease in the medium. More than 1,000 colonies per deletion site were screened for activity by plate assay. More than 30 colonies per site were further characterized by sequencing, SDS-PAGE immunoblots, and solution activity assays. These analyses showed that deletions of more than five amino acids at any site abolished protease activity

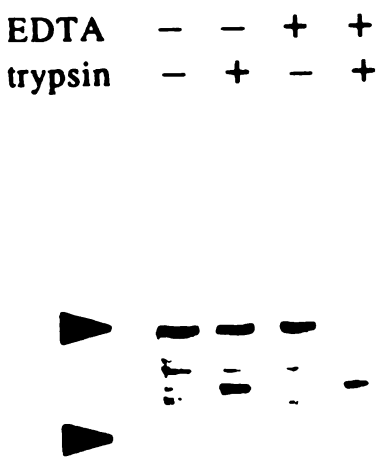


Figure 2.7. **Accessibility to trypsin.** Sensitivity of the wild type precursor to trypsin was tested for whole cells and EDTA-permeabilized cells as indicated. The original state of degradation by endogenous proteases is indicated in the first and third lanes (whole cells and EDTA-permeabilized cells, respectively), and their sensitivity to exogenously added trypsin is shown in the second lane (whole cells) and fourth lane (EDTA permeabilized cells). The position of the precursor is indicated by the arrow.

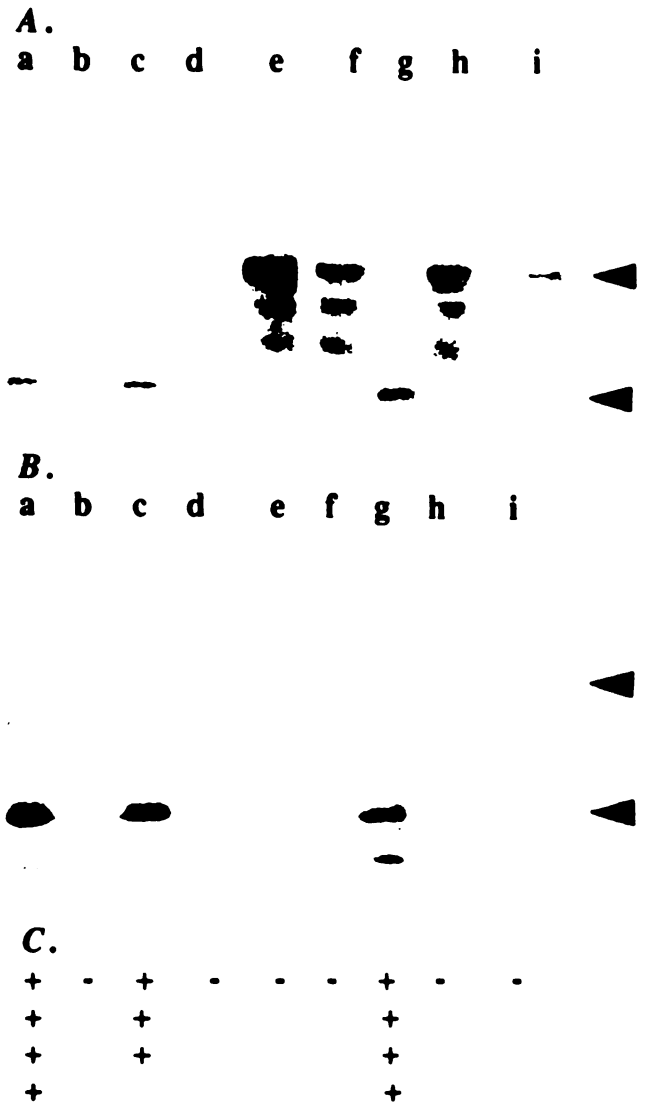


Figure 2.8. **Correlation of activity and secretion.** Representative immunoblots of cell extracts and dialyzed media from pro region deletion mutants. The arrows indicate the positions of the precursor and the mature protease. *A:* cells: lane a, pALP6 (wild type). lane b, uninduced pALP6 . lane c, N-terminal deletion of 3 amino acids. lane d, N-terminal deletion of 6 amino acids. lane e, 3 amino acid deletion at the *Bgl* I site. lane f, 13 amino acid deletion at the *Hind* III site. lane g, one amino acid deletion at the *Dra* I site. lane h, 10 amino acid deletion at the *Dra* I site. lane i, 4 amino acid deletion at the *Nar* I. *B*, media, samples as above. *C.* activity observed in the media.

in both the cells and the medium. Furthermore, deletions within the pro region that produced inactive protease invariably disturbed secretion to the medium and vice versa, despite the fact that the protease-coding segment itself was not mutated. One explanation of these data is that the folding and secretion functions of the pro region may not be separable. Another interpretation is that secretion requires not only an intact pro region, but protease activity as well.

Accordingly, we investigated whether enzymatic activity was required for export across the outer membrane, or whether proper folding was sufficient. In order to produce an inactive but correctly-folded protease, we constructed the complementation vector pCOMP4-SA195, which supplies a wild type copy of the pro region *in trans* to the protease region containing the active site mutation. Previous results had shown that the pro region can supply the necessary folding information to the protease region *in trans* (Silen and Agard, 1989). Such a construct should allow proper folding of the protease region to occur without producing the active protease, and precludes the necessity for auto-proteolytic processing between the pro and protease regions. In contrast to both the unprocessed form of this mutant and to the misfolded Δ-pro-α-lytic protease, the folded but inactive mutant protease was found both in the medium and in the cells (Figure 2.9A). It is important to note that this fractionation of protease between cells and medium is seen in all constructs expressing functional α-lytic protease (either the intact protein or via complementation) in *E. coli*. Over a period of days, α-lytic protease continues to accumulate in the medium, and the cell-associated fraction remains constant or decreases. In these cases, the cell-associated form is mature sized, shows α-lytic protease activity, and is presumed to be a soluble periplasmic species.

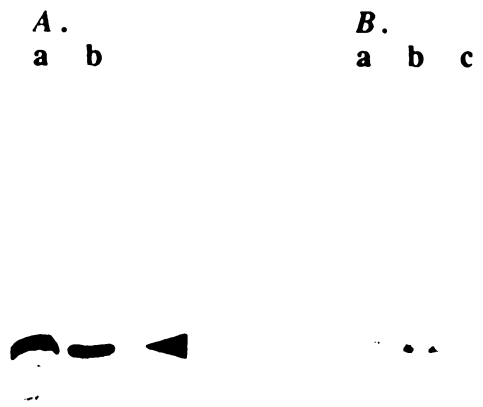


Figure 2.9*A*. **Complementation of the folding defect in the active site mutant.** *A*, pCOMP4-SA195 was expressed at 22°C in *E.coli* MH1 cells (see Materials and Methods for conditions). The media and cells were separated by centrifugation and equivalent amounts were loaded on a discontinuous SDS gel. The arrow indicates the migration position of mature.α-lytic protease.The immunoblot shows that the mutant is of mature size and is found in the cells (lane a) as well as the media (lane b). *B*, the supernatants of pALP5 and pCOMP4-SA195 were each dialysed against distilled water, lyophilized and brought up in a small volume of distilled water. The samples were treated briefly with trypsin to remove bound pro region and then combined with low pH sample buffer and loaded on a non denaturing, non-reducing, low pH gel . The immunoblot of the gel is shown above. *Lane a*, wild type α-lytic protease. *Lane b*, pCOMP4-SA195 treated with trypsin. *Lane c*, trypsin.

The inactive mutant protease secreted by the complementation system comigrates with wild type α-lytic protease on a nondenaturing native gel (Figure 2.9B), verifying that the two have substantially similar tertiary structure. Because the pro region is known to bind tightly to the protease region until it is proteolyzed (Baker et al., 1992), the fractionated samples containing inactive mutant protease (pCOMP+SA195) were trypsinized briefly before electrophoresis. The fact that the mutant protease region is trypsin resistant provides additional evidence that it is compactly folded.

Immunoblots of cell extracts and media containing wild type α-lytic protease have revealed that cleavage of the pro region occurs prior to secretion into the medium. Furthermore, when the same blots are probed with antibodies directed against the pro region, it can be seen that the pro region itself is released into the medium. Nevertheless, when expressed alone in the pPRO2 construct, the pro region remains cell associated (Fig. 2.10). Attempts at further fractionation have been hampered by the apparent instability of the pro region when expressed alone.

a b c d

Figure 2.10. **Location of the pro region.** Immunoblots of pPRO2 and pALP5, using a primary antibody directed against a pro region fusion protein (see materials and methods). The pro region content of the cells and the media were compared under conditions where the pro region is produced alone (pPRO2) versus in the wild type construct where it is produced with mature protease (pALP5). Pro region content of pPRO2-bearing cells (lane a) and pPRO2 medium (lane b), pALP5-bearing cells (lane c) and medium (lane d).

Discussion

In this paper, we have sought to identify factors that interfere with the extracellular secretion of α-lytic protease. It has been established that the pro region plays a crucial role in the folding of the protease region either *in vivo* (Silen et al., 1989; Silen and Agard, 1989) or *in vitro* (Baker et al., 1992). The pro region must be both present and cleaved to allow α-lytic protease to be secreted into the medium. In all other cases, such as the Δ-pro-α-lytic protease or the precursor produced by the SA195 mutation, we find that the α-lytic molecule becomes tightly associated with the outer membrane in a manner that cannot be released by detergents, high salt or high pH, but can be released by treatment with protein denaturants. The precursor is more sensitive to proteolysis by added trypsin in EDTA-permeabilized cells than in whole cells. The possibility existed that these proteins had aggregated and associated with the outer membrane either *in vivo* or during cell lysis. The appearance of a protected band upon trypsinization argues against the possibility that the precursor exists as an aggregate. Wehave also tried to address the problem of aggregates sticking to the outer membrane by using flotation, by varying the extraction and gradient conditions, and by testing the solubility properties of all cell-associated species in Triton X-100 with EDTA. The results support a *bona fide* association of the precursors and Δ-pro-α-lytic protease with the outer membrane, rather than an artifactual trapping of aggregates. Beyond this, the extraction properties cited above are suggestive of a molecule arrested in its interaction with protein docking or translocation machinery, or tightly bound to an outer membrane protein in a non-translocation-competent fashion.

What is the role of the pro region in the secretion process? One possibility had been that the pro region carries the "targeting signal" that mediates association with the outer membrane. This does not appear to be the case with α-lytic protease. The fact that Δ-pro-α- lytic protease is also tightly associated with the outer membrane indicates that, although the pro region is required to complete transport across the outer membrane, it is not required for association with the membrane. Another possibility, as suggested for the IgA protease (Pohlner et al., 1987), is that the pro region actually forms the channel through which the protease domain passes. For wild type α-lytic protease, the pro region is cleaved in the periplasmic space and accumulates in the medium. When expressed *in trans* to the protease region, it also appears in the medium. When expressed alone, however, the pro region accumulates in the cells but not in the medium. This data must be interpreted with some caution, as it is probable that the pro region is unstable in the absence of the α-lytic protease region. Nevertheless, for the wild type construct, there is an apparent interdependence for secretion of the pro region and the protease region. Furthermore, Karen Smith's work demonstrates a strong correlation between secretion and folding of the protease region in mutants with disrupted pro regions. Finally, the correctly folded but inactive version of α-lytic protease (pCOMP4-SA195) is efficiently transported. Taken together, these data suggest that the primary function of the pro region is to effect proper folding of the protease region, and that only properly folded molecules are efficiently transported across the outer membrane. It is not yet possible in our system to distinguish between selective transport of folded proteins (comparable to eukaryotic nuclear import, Dingwall and Laskey, 1986) and selective retention of misfolded proteins (comparable to BiP-mediated

retention of misfolded proteins in the endoplasmic reticulum, Gething et al., 1986).

α-Lytic protease is one of several proenzymes that have been shown to possess pro region-dependent export characteristics when expressed in *E. coli* (Miyazaki et al., 1989; Silen et al., 1989; Terada et al., 1990). The conditions for export appear to be protease specific and to correlate with the conditions for maturation to the active form, which vary widely for the different proteases. For example, aqualysin we requires temperatures of 65°C for final processing and release from the outer membrane (Terada et al., 1990), while α-lytic protease is efficiently processed and secreted only at temperatures below 30°C (Silen et al., 1989). Other evidence for the involvement of folding in targeting to or across the outer membrane comes from studies of the Omp F porin. Sen and Nikaido (1990) have successfully trimerized Omp F following incorporation of folded monomers into cell envelope preparations in the presence of small amounts of Triton-X-100. Their work suggests that the outer membrane can accept prefolded molecules. In contrast, the inner membrane appears to require linear insertion of proteins.

In the best characterized cases, it appears that targeting to various eukaryotic and prokaryotic cellular locations is mediated by primary structural cues (e.g., signal sequences (Blobel and Dobberstein, 1975) or specific targeting signals such as KDEL for ER retention (Munro and Pelham, 1987), or PKKKRKV for nuclear import of SV40 large T antigen (Kalderon et al., 1984)). Many attempts have been made to discover a primary sequence responsible for targeting proteins into or across the outer membrane of Gram-negative bacteria (Nikaido and Vaara, 1987). Because no such signal or independently transporting domain has yet been discovered, it has been suggested that insertion into or transport across the outer membrane may be

encoded in a tertiary structural cue (Sen and Nikaido, 1990; Model and Russel, 1990; Dornmair et al., 1990).

We propose a model (Figure 2.11) for transport of proteins across the outer membrane of *E.coli*, wherein tertiary and quaternary structure play a significant role. However, our data suggest that while completion of translocation or release from the outer membrane requires that the protein be properly folded (a tertiary cue), association with the outer membrane and perhaps initiation of transport does not require a tertiary signal (misfolded proteins associate very strongly). In this model, the role of pro regions can be understood in that they are necessary for assisting in the folding of the protein to be secreted, and not directly involved in the transport process *per se*. Furthermore, this model proposes that outer membrane translocation involves mechanisms very different from those elucidated for transport across the inner membrane of *E.coli*, mitochondrial or chloroplast membranes, or the ER membrane,where a targeting sequence and the absence of tertiary structure are the major requirements (reviewed by Eilers and Schatz, 1988). Further studies are required to determine whether the folded state is being recognized for transport or the incorrectly folded state is being recognized for retention.

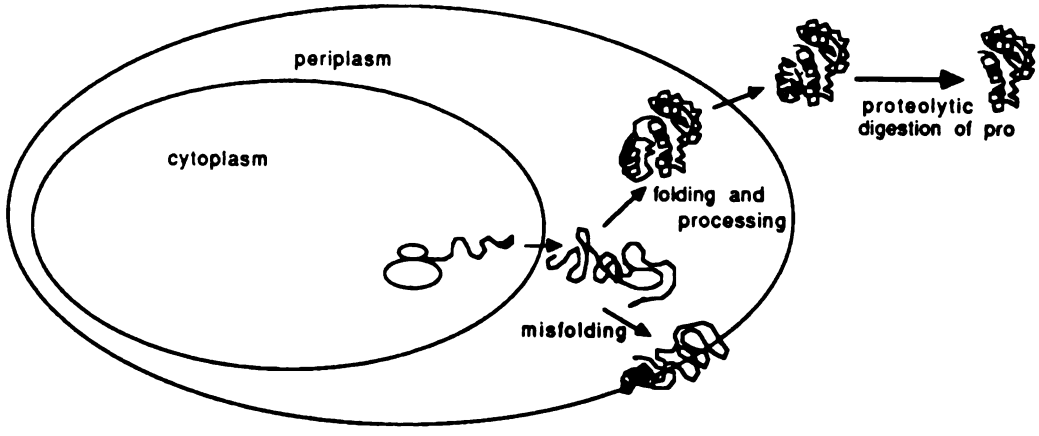


Figure 2.11. **Maturation of α-lytic protease.** The preproenzyme is transported across the inner membrane to the periplasmic space, where the signal (pre) sequence is removed. Upon proper folding, the precursor is cleaved between the pro region and the protease region. α-lytic protease activity is required for the cleavage. The two regions have a high affinity for one another, and therefore remain as a complex while they are conveyed across the outer membrane to the media. Over a period of time, the pro region is further degraded, leaving the mature protease. If the protein does not fold correctly or cleavage between the pro region and the protease region does not occur, the misfolded/precursor form becomes tightly associated with the outer membrane.

Acknowledgements

We extend many thanks to Dick Shand for assistance in looking for inclusion bodies and for useful discussions. We also thank John Reidhaar-Olson and Peter Walter for critical reading of the manuscript. We are grateful to F. Szoka for the use of his scintillation counter, and to R.B. Kelly for the use of his refractometer. Funding for this research was provided by the Howard Hughes Medical Institute.

References for Chapter Two: Localization Studies

Bachovchin,W.W., Wong, W.Y.L., Farr-Jones, S. Shenvi, A.B., and C.A.Kettner. 1988. Nitrogen-15 NMR spectroscopy of the catalytic-triad histidine of a serine protease in peptide boronic acid inhibitor complexes. *Biochemistry* 27:7689-7697

Baker, D., Silen, J.L., and D.A. Agard. 1992. A protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins* 12:339-344

Birnboim, H.C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* 7:1513-1523

Blobel, G. and B.Dobberstein. 1975. Transfer of proteins across membranes. *J. Cell Biol.* 67:835-851

Bone, R.B., Frank, D., Kettner, C.A., and D.A.Agard. 1989a. Structural analysis of specificity: α-lytic protease complexes with analogues of reaction intermediates. *Biochemistry* 28:7600-7609

Bone, R.B., Fujishige, A., Kettner, C.A. and D.A. Agard. 1991. Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry:* 30:10388-10398

Bone, R.B., Shenvi, A.B., Kettner, C.A., and D.A.Agard. 1987. Serine protease mechanism: Structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. *Biochemistry* 26:7609-7614

Bone, R.B., Silen, J.L., and D.A.Agard. 1989b. Structural plasticity broadens the specificity of an engineered protease. *Nature* 339:191-195

Dingwall, C. and R.A.Laskey. 1986. Protein import into the cell nucleus. *Ann. Rev. Cell Biol.* 2:367-390

Dornmair, K., Kiefer, H., and F.Jähnig. 1990. Refolding of an integral membrane protein. Omp A of *Escherichia coli. J. Biol. Chem.* 265:18907-18911

Eilers, M. and G.Schatz. 1986. Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature* 322:228-232

Eilers, M. and G.Schatz. 1988. Protein unfolding and the energetics of protein translocation across biological membranes. *Cell* 52:481-483

Filloux, A., Bally, M., Ball, G., Akrim, M., Tomassen, J., and A.Lazdunski. 1990. Protein secretion in Gram-negative bacteria: transport across the outer membrane involves common mechanisms in different bacteria. *EMBO J.* 9:4323-4329

Gething, M.-J., McCammon, K., and J. Sambrook. 1986. Expression of wild-type and mutant forms of Influenza hemagglutinin: The role of folding in intracellular transport. *Cell* 46:939-950

Hall, M.N., Hereford, L., and I. Herskowitz. 1984. Targeting of E.coli beta-galactosidase to the nucleus in yeast. *Cell* 36:1057-1065

Hall, M.N., Schwatrz, M., and T.J.Silhavy. 1982. Sequence information within the *lamB* gene is required for proper routing f the bacteriophage λ receptor protein to the outer membrane of *Escherichia coli* K-12. *J. Mol. Biol.* 156:93-112

Hames, B.D. 1981. An introduction to polyacrylamide gel electrophoresis. *In* Gel Electrophoresis of Proteins. Hames, B.D. and D.Rickwood, eds. IRL Press, Oxford. 1-91.

Hattori, M., and Y.Sakaki. 1986. Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* 152:232-238

Hindennach, I. and U. Henning. 1975. The major proteins of the *Escherichia coli* outer cell envelope membrane. *Eur. J. Biochem.* 59:207-213

Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R., and J.H.Richards. 1973. Carbon nuclear magnetic resonance studies of the histidine residue in α -lytic protease. *Biochemistry* 12:4732-4743

James, M.N.G., Del Baere, L.T.J., and G.D.Brayer. 1978. Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. *Can. J. Biochem.* 56:396-402

Kalderon, D., Roberts, B.L., Richardson, W.D., and A.E.Smith. 1984. A short amino acid sequence able to specify nuclear location. *Cell* 39:499-509

Karkhanis, Y.D., Zeltner, J.Y., Jackson, J.J., and D.J.Carlo. 1978. A new and improved microassay to determine 2-keto-3-deoxyoctonate in lipopolysaccharide of Gram-negative bacteria. *Anal. Biochem.* 85:595-601

Kunkel, T.A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U.S.A.* 82:488-492

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685

Mackman, N., Nicaud, J.M., Gray, L., and I.B.Holland. 1985. Identification of polypeptides required for the export of

haemolysin 2001 from *E. coli.*. *Mol. Gen. Genet.* 201:529-536

Miyazaki, H., Yanagida, N., Horinouchi, S., and T.Beppu. 1989. Characterization of the precursor of *Serratia marcescens* serine protease and COOH-terminal processing of the precursor during its excretion through the outer membrane of *Escherichia coli. J. Bact.* 171:6566-6572

Mizushima, S. and H.Yamada. 1975. Isolation and characterization of two outer membrane preparations from *Escherichia coli. Biochim. Biophys. Acta* 375: 44-53

Model, P. and M.Russel. 1990. Prokaryotic secretion. *Cell* 61:739-741

Munro, S. and H.R.B.Pelham. 1987. A C-teminal signal prevents secretion of luminal ER proteins. *Cell* 48:899-907

Neidhardt, F.C., Bloch, P. L., and D.F.Smith. 1974. Culture medium for enterobacteria. *J.Bact.* 119:736-747

Neu, H.C. and L.A.Heppel. 1965. The release of enzymes form *Escherichia coli* by osmotic shock during the formation of spheroplasts. *J. Biol. Chem.* 240:3685-3692

Nikaido, H. and M.Vaara. 1987. Outer membrane. *In Escherichia coli* and *Salmonella typhimurium* . F.C.Neidhardt, ed. in chief. Am. Soc. Micro., Washington, DC. 7-22.

Pohlner, J., Halter, R., Beyreuther, K., and T.F.Meyer. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* 325:458-462

Pugsley, A. P. 1989. Protein Targeting. San Diego, Academic Press

Pugsley, A. P., d'Enfert, C., Reyss, I., and M.G.Kornacker. 1990. Genetics of extracellular protein secretion by Gram-

Randall, L. L., and S. J. S.Hardy. 1986. Correlation of competence

for export with lack of tertiary structure of the mature species: A study *in vivo* of maltose-binding protein in *E.coli. Cell* 46:921-928

Schnaitman, C.A. 1974. Outer membrane proteins of *Escherichia coli.* IV. Differences in outer membrane proteins due to strain and cultural differences. *J. Bact.*118:454-464

Sen, K. and H.Nikaido. 1990. *In vitro* trimerization of OmpF porin secreted by spheroplasts of *Escherichia coli. Proc. Natl. Acad. Sci.* 87:743-747

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., Frank, D., Fujishige, A., Bone, R., and D.A. Agard. 1989. Analysis of prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro region is required for activity. *J. Bact.*171:1320-1325

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

Tabor, S., and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci.* 84:4767-4771

Terada, I., Kwon, S., Miyata, Y., Matsuzawa, H., and T.Ohta. 1990. Unique precursor structure of an extracellular protease, aqualysin I, with NH_2- and COOH-terminal pro-sequences and its processing in *Escherichia coli. J. Biol. Chem.* 265: 6576-6581

Toma, S., Campagnoli, S., De Gregoriis, E., Gianna, R., Margarit, I., Zamai, M., and G. Grandi. 1989. Effect of Glu-

143 and His-231 substitutions on the catalytic activity and secretion of *Bacillus subtilis* neutral protease. *Protein Eng.* 2:359-364

Wandersman, C., and P.Delepelaire. 1990. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc. Natl. Acad. Sci.* 87:4776-4780

Whitaker, D.R. 1970. The α -lytic protease of a myxobacterium. *In* Meth. Enz. 19:599-613

Footnotes

1. Abbreviations used in this text: EDTA, ethylenediaminetetraacetic acid disodium salt; LB, Luria-Bertoni; MOPS, morpholinepropanesulphonic acid; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

Chapter 3: **Kinetics of Export**

During my graduate career, I was fortunate in being able to attend a number of excellent scientific conferences. At one such conference, I met Professor L. Randall, with whom I discussed the work that is now published and was included here as Chapter Two. While Dr. Randall's work concerns transport across the inner membrane of E. coli, I found it useful to follow her work in order to establish guidelines and standards for the development of our own work on the outer membrane. As a guest in Dr. Randall's lab during a memorable two weeks of November of 1991, I learned most of the techniques employed in the following paper.

Dr. Agard returned from another conference at this time full of ideas and enthusiasm for the project that I had undertaken of determining the kinetics of export of α -lytic protease. Indeed, it was his enthusiasm that convinced me to follow up our original experiments in *E. coli* with similar ones in the native host for α-lytic protease, *Lysobacter enzymogenes.* The result is a tantalizing suggestion that there are accessory proteins for outer membrane transport. Current lab member Barry Wilk and coming postdoctoral fellow Alan Derman will pursue the project further. The following chapter has been submitted as an article to the Journal of Bacteriology.

Kinetics of Extracellular Secretion of α-Lytic Protease

Amy Fujishige-Boggs, Barry J. Wilk and David A. Agard

Howard Hughes Medical Institute and the

Departments of Pharmaceutical Chemistry and Biochemistry

University of California at San Francisco

San Francisco, CA 94143-0448

Abstract

α-Lytic protease is an extracellular serine protease secreted by the Gram-negative soil bacterium, *Lysobacter enzymogenes* (Whitaker, 1970). Synthesized as a preproenzyme (Silen et al., 1988), the 166 residue pro region is required transiently to facilitate the folding of the 198 residue protease domain (Silen et al., 1989; Silen and Agard, 1989, Baker et al 1992a,b). Previous studies using *E. coli* have demonstrated a clear link between folding and extracellular secretion: only correctly folded molecules are secreted (Fujishige et al, 1992; Silen et al., 1989). As a first step in dissecting the mechanisms of extracellular secretion in *E. coli.*, we demonstrate here that the protease passes through the periplasm en route to the media. In addition, we show that the rate of export of the protease expressed in *E. coli* is quite slow ($t_{1/2}$~4 h). By contrast, the half-time of export in the native host is less than 5 minutes, comparable to export of other proteins from Gram-negative bacteria (Pugsley, 1992; Hirst and Holmgren, 1987; Wong et al., 1989; He et al., 1991). Extracellular secretion of heterologous proteins in *E. coli* in the absence of accessory genes is quite unusual. Although slow, the secretion of α-lytic protease in the absence of accessory proteins suggests that a primitive or residual two-step export system is present in *E. coli.* It is also probable that the native host provides accessory proteins such as those found in other Gram-negative bacteria. In the heterologous system described here, such proteins would facilitate, but not be essential for export.

Introduction

Transport of proteins across the outer membrane of Gram-negative bacteria is a complex and important process. Exposure of adhesion molecules (Krogfelt, 1991) and extracellular secretion of toxins (Berkowitz, 1991) and other proteins (Galyov et al., 1993) are integral steps in pathogenesis. In addition, extracellular secretion could be of enormous use in the laboratory and in biotechnology, where the Gram-negative bacterium *E. coli* remains a popular expression host. Curiously, little is known about the targeting signals or the mechanisms of extracellular transport. Existing data suggests that there may be significant differences between the mechanisms of transport across the cytoplasmic and outer membranes. There is now evidence that the targeting signal for extracellular secretion is three dimensional (Hirst and Homgren, 1987; Sen and Nikaido, 1990) rather than a linear signal such as is used by cytoplasmic membrane transport systems. In addition, recent studies with α-lytic protease and pullulanase demonstrate that, unlike translocation across the cytoplasmic membrane, which requires that proteins be maintained in an unfolded conformation (Randall and Hardy, 1986; Eilers and Schatz, 1986), secretion through the outer membrane requires that proteins be properly folded (Fujishige et al., 1992; Pugsley 1993).

There appear to be at least two different mechanisms by which proteins can exit the Gram-negative cell. Several toxins and proteases appear to cross the inner and outer membranes in a single step (Mackman et al., 1985; Filloux et al., 1990; Wandersman and Delepelaire, 1990) at Bayer junctions, which are areas of close apposition between the two membranes. Other proteins have been shown to cross the two membranes separately in what is called the two-step mechanism. In this mechanism, proteins are

first translocated across the cytoplasmic membrane in a signal sequence- and Sec-dependent manner, and only subsequenly translocated across the outer membrane. This is also referred to as the extended general secretory pathway (Pugsley, 1993).

Of the several two-step protein export systems that have been characterized, all have been shown to be dependent on accessory proteins. In the case of pullulanase, where the export system has been successfully reconstituted in *E. coli*, passage across the outer membrane is dependent upon the expression of the *pul* operon accessory proteins. There is significant homology between proteins of the *exe, xcp, out* , *pul* and other operons, which encode accessory transport proteins from many different Gram negative bacteria (Filloux et al., 1990; for review, see Pugsley, 1993). Thus, it appears that there will be strong mechanistic similarities in these transport systems. Proteins homologous to these transport proteins have not been identified in *E. coli.*

In contrast to the pullulanase system, heterologous expression of prepro-α-lytic protease alone is sufficient for its extracellular transport from *E. coli.* α-Lytic protease is an extracellular serine protease from the Gram-negative soil bacterium, *Lysobacter enzymogenes* (Whitaker, 1970). The 166 residue pro region has been shown to be required both *in vivo* (Silen et al., 1989; Silen and Agard, 1989) and *in vitro* (Baker et al., 1992a, b) for the correct folding of the 198 residue protease domain. The observed coupling of folding and secretion (Fujishige et al., 1992) and the ability to control the folding of α-lytic protease via manipulation of the pro region make this an ideal system for probing extracellular secretion in *E. coli.* As a first step in this characterization, we have determined whether α-lytic protease is secreted via the one-step or two-step pathways. A key difference

in these two mechanisms is that proteins utilizing the two-step mechanism reside transiently in the periplasm. Below we present pulse-chase data to confirm that α-lytic protease uses the two-step mechanism of export when heterologously expressed in *E. coli*. In addition, these experiments allowed us to examine the kinetics of export across the outer membrane.

While export of α-lytic protease from *E. coli* occurs in the absence of exogenous factors, we considered the possibility that the native host of α-lytic protease might contain accessory proteins that are not essential but which make the extracellular transport process more efficient. A comparison of the rates of export of α-lytic protease in *E. coli* and *L. enzymogenes* shows that this is likely to be the case.

Materials and Methods

Media, Strains, and Plasmids-- *L. enzymogenes* 495 (ATCC29487) was the gift of F.D. Cook. Cultures were grown in M9 media (Maniatis et al., 1982) supplemented with 0.15% casamino acids (Difco Laboratories, Detroit, MI) and subsequently transferred to M9 media supplemented with all amino acids except cysteine and methionine. All *L. enzymogenes* cultures were grown in the presence of 40 μg/ml kanamycin, to which *L. enzymogenes* is resistant (Horne, 1987), in order to reduce possible contamination.

Plasmid pALP5 is a pBR322 derivative that contains the gene for pro-α-lytic protease fused to the signal sequence of alkaline phosphatase and driven by the PhoA promoter, as has been described elsewhere (Silen et al., 1989). Plasmids were prepared by the alkaline lysis method of Birnboim and Doly (1979). *E. coli* MH1 cells (*araD139 ΔlacX74 galU galK hsrrpsL;* Hall et al., 1984) were transformed with pALP5 and grown first in high phosphate minimal media and subsequently transferred to low phosphate minimal media as described by Tommassen and Lugtenberg (1980). All pALP5-bearing *E. coli* cultures were grown in the presence of 100 μ g/ml carbenecillin.

Plasmid pALP12 is a derivative of pUHE21-2 that contains the gene for α-lytic protease fused to the signal sequence of alkaline phosphatase and driven by an IPTG-inducible T7 P_{A1} promoter. The pUHE21-2 vector was made in the laboratory of H. Bujard in Heidelberg, and contains a synthetic T7 P_{A1} early promoter combined with two lac operators inserted into plasmid (Deuschle et al., 1986). This vector was modified in our laboratory to contain the alkaline phosphatase signal sequence (Chang et al., 1986) using a cassette comprised of a complementary set of 65 base long oligonucleotides

(Howard Hughes Medical Institute DNA Facility, UCSF, San Francisco, CA) with BamH1 and Xba1 overhangs at the 5' and 3' ends, respectively. When ligated into the BamH1/Xba1 sites of pUHE21-2, the cassette places the signal sequence two amino acids downstream of the start methionine. The cassette also contain a unique restriction site, Eag1, at the end of the signal sequence and twelve bases upstream of the Xba1 site. The α-lytic protease gene was excised from pALP5 using the Eag1 site (between the signal sequence and the mature region) and the Xba1 site (at the end of the gene), and ligated into the modified pUHE21-2 vector using the same sites, to produce pALP12.

E. coli D1210 cells (*F lacI^{Q+}lacO gal pro leu thi end hsm hsr recA rpsL*; Sadler et al., 1980) were transformed with pALP12 and grown in LB broth supplemented with 20mM $MgCl_2$ and 100 mg/ml carbenecillin. Cells were induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when the optical density reached 1.2 at 600 nm.

Because expression of α-lytic protease is temperature-sensitive in *E. coli*, with a transition temperature near 30°C and optimal expression at or below 22°C (Silen et al., 1989; Fujishige-Boggs et al., manuscript in preparation), and because *Lysobacter enzymogenes* grows optimally at or below 25°C, all cultures were grown and labelled at 22° C, except pALP12-bearing D1210 cells, which were grown at 18°C to reduce lysis.

Time course of Expression and Secretion-- Samples were taken from induced pALP12 cultures at intervals for five days. The samples were assayed for optical density at 600 nm and for α-lytic activity as previously described (Silen et al., 1989). In addition, the cells and the media were separated by centrifugation and analyzed by immunoblots of SDS-gels (see below). Gel contents were transferred to nitrocellulose and probed with affinity

purified rabbit anti-α-lytic protease antibody as previously described (Fujishige et al., 1992) followed by alkaline phosphatase-conjugated goat-anti-rabbit antibody (Bio-Rad Laboratories, Richmond, CA). Bands were visualized using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl-phosphate (Bio-Rad Laboratories, Richmond, CA).

Labelling Conditions and Fractionation-- pALP5 cultures were grown to mid-log phase and labelled for 15 sec to 5 min (as indicated) with ^{35}S-methionine (NEN, >800 Ci/mmol). Labelling was terminated by the addition of >500 fold excess cold methionine. For experiments lasting longer than one hour, media was changed immediately after the cold methionine was added (and the first sample removed) to LB Broth supplemented with 1mM cold methionine and the appropriate antibiotic. Samples were taken at intervals indicated and placed immediately in a prechilled beaker in an ice-water slurry. After one minute, the samples were transferred to eppendorf tubes and the cells and media were separated by 40 second spin at 4°C (Eppendorf centrifuge, model 5415). The supernatant was removed and respun at 4°C for 15 min. The supernatant of this second spin will be referred to as the media fraction.

E. coli cell pellets were fractionated to obtain a native periplasmic portion as follows: Pellets were resuspended in 0.25ml 0.1M Tris, 0.5mM EDTA, 0.5M sucrose (pH 8.2). 20 μl of 2 mg/ml lysozyme (freshly prepared) was added to the resuspended pellets, followed immediately by 0.25 ml ice cold distilled water. The samples were placed on ice for 5 min after which 10μl 1M $MgSO_4$ was added. The samples were centrifuged in eppendorf tubes at 4°C for 40 sec to separate the periplasmic contents (supernatant) from the remaining cell contents (pellet) (Randall and Hardy, 1984).

L. enzymogenes is unusually difficult to lyse. Cycles of freezing and thawing, and/or treatment with lysozyme do not give satisfactory results (data not shown). Therefore, a combination of lysozyme treatment and secretion was employed in order to gather the labelled cell contents. After removal of the media, the cell pellets of *L. enzymogenes* were treated with 1 mg/ml lysozyme for one hour at room temperature, and centrifuged for 5 minutes at 4°C. The supernatant was saved (see below), and the treated cell pellets were placed in fresh media to secrete completely their contents under typical growth conditions for three hours. Exported cell contents were gathered as the supernatants of 5 minute spins at 4°C. Cell content of labelled α-lytic protease was determined from immunoprecipitation (see below) of the combined exported cell contents (after removal of original medium for that time point) and lysozyme-treated cell supernatants.

Immunoprecipitation and SDS-PAGE-- Protein A sepharose beads (Pierce) were combined with anti-α-lytic protease antibodies (Silen et al., 1989) for >2h on ice. After a 4 second centrifugation to settle the beads, the supernatants were removed and the beads were washed with 10mM Tris-acetate, pH 7.6 to remove unbound antibody, leaving the bound antibody-beads as a 1:1 slurry. Samples (media and cell contents, or media and periplasmic extracts) were immunoprecipitated overnight at 4°C with 20 μ l of the 1:1 bead slurry. Immunoprecipitates were washed in 0.5 ml 0.15M NaCl, 10mM Tris-acetate pH 7.6, 5mM EDTA, 0.5% TritonX-100, 0.1% SDS, and rinsed twice with 10 mM Tris-acetate pH 7.6. Immunoprecipitated proteins were recovered by adding 20 μ l 3x Laemmli sample buffer (Laemmli, 1970) directly to the beads and boiling for 5 min. Total labelled proteins and immunoprecipitated proteins were analyzed by SDS-PAGE (Laemmli, 1970)

using 5%/12% discontinuous acrylamide-bisacrylamide gels. Proteins were stained with Coommassie brilliant blue R-250 (Sigma, St. Louis , Mo.) and treated with Amplify™ (Amersham Inc., Arlington Heights, IL) to enhance signal.

Autoradiography and Phosphorimaging-- Preflashed Amersham film(Amersham Inc, Arlington Heights, IL) was exposed to dried gels at -80°C and developed using an automated film processor (AFP Imaging Corp., Elmsford, N.Y.). In addition, a phosphorimaging screen (Molecular Dynamics, Sunnyvale, CA) was exposed to gels. Scanning and quantitation were performed using Imagequant™ (Molecular Dynamics) software.

Results

A time course of α-lytic protease expression and secretion in *E. coli* is shown in Figure 3.1. Since α-lytic protease is synthesized with a signal sequence and the protease domain has three disulfide bonds, the observed cellular activity undoubtedly represents active enzyme in the periplasmic space. The extremely slow secretion into the media seen in this experiment is highly characteristic of all our expression data for α-lytic protease. The temporal behavior is essentially independent of promoter, temperature (within the permissive range) or *E. coli* host strain, although the actual enzyme levels are sensitive to these parameters. Western blot analysis demonstrated that the observed enzyme activity levels are indicative of actual protein levels (data not shown). Optical density measures indicate that the pALP12-bearing D1210 cultures reach stationary phase at roughly 40 hours under the growth conditions, and that the cells did not lyse appreciably during the course of the experiment (data not shown). The delayed rise of activity in the media is also characteristic of α-lytic expression from the native host, *Lysobacter enzymogenes* (S. Farr-Jones and J. Davis, pers., comm.) where secreted activity peaks between 24 and 48 hours (expression temperatures ~27°C), suggesting that this pattern is not an artifact of heterologous expression in *E. coli*. Because periplasmic levels of protease remain roughly constant while the enzyme accumulates in the media, it was unclear whether the periplasmic molecules were export-competent intermediates, or whether they were a small population of aberrently targeted molecules.

Our initial goal was to determine whether α-lytic protease uses the one-step or the two-step mechanism for extracellular secretion in *E. coli*.

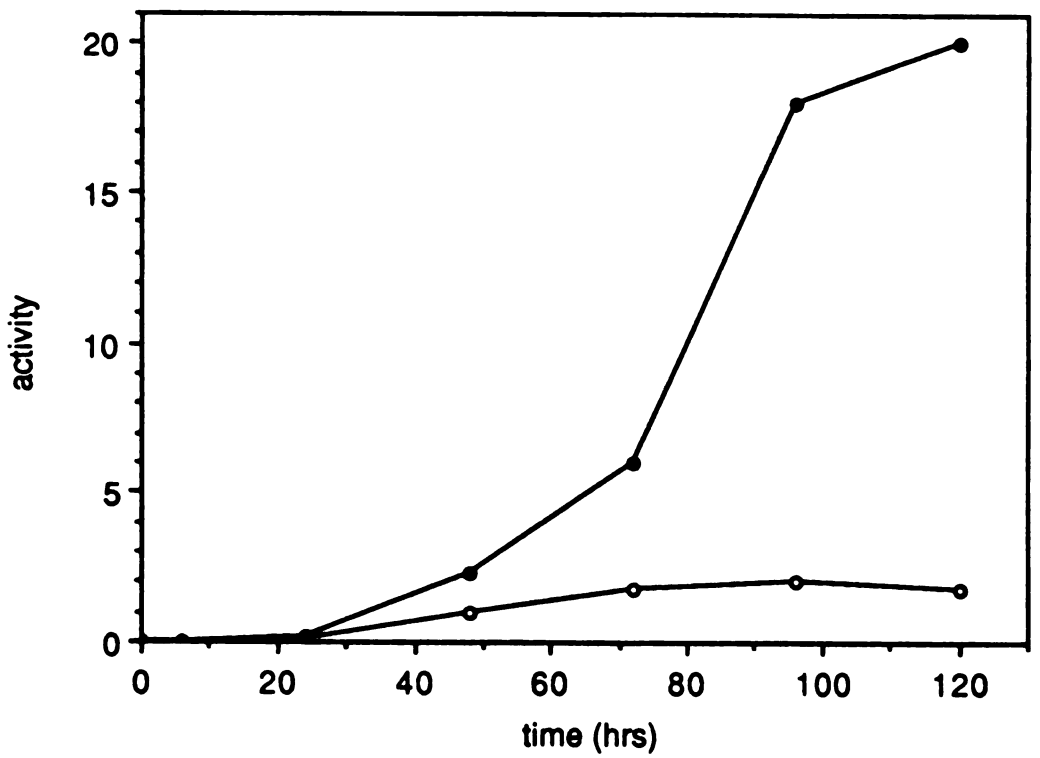


Figure 3.1. Typical time course of expression of α-lytic protease. pALP12-bearing *E. coli* D1210 cells were induced (t=0) at an optical density of 1.2 at 600 nm. Samples were taken daily. Active α-lytic protease is first observed in the cellular fraction (open circles) and is later observed in the media (closed circles). The activity in the cells levels off as the culture reaches stationary phase (see text), while activity in the media increases dramatically between 40 and 100 hours.

Since the two-step mechanism requires transient residence in the periplasmic space en route to the media, we sought to distinguish the two mechanisms by pulse-chase experiments combined with cell fractionation. Pilot studies revealed that immunoprecipitation of denatured α-lytic protease was at least 50-fold less efficient than immunoprecipitation of native protein (data not shown). For this reason, it was difficult to examine SDS-solubilized membrane fractions. Nevertheless, immunoprecipitation of the periplasmic and the media fractions was practical under non-denaturing conditions.

Initial studies revealed that α-lytic protease export from *E. coli* was extremely slow, taking several hours (data not shown). This allowed us to increase the labeling period to 5 minutes, resulting in a better signal. In addition, because the chase period lasted for 11 hours, it was important to change the media after the first time point in order to avoid continuous labeling. Samples from all time points were fractionated to obtain media and periplasmic components. Analysis of these fractions revealed that labeled α-lytic protease does indeed arrive in the periplasm, and that this population chases to the medium (Figure 3.2).

As had been indicated by the initial studies, the kinetics proved to be quite slow, with a half-time of approximately 4 hours (Figure 3.3). Cell counts by plating show that more than 80% of the original number of cells are viable at this time (data not shown), indicating that the protein present in the media has been exported, rather than released by lysis. Although the experimental data from other systems are not extensive, previous analyses of outer membrane transport kinetics indicate a $t_{1/2}$ ranging between 1 and 15 minutes (Pugsley, 1992; Hirst and Holmgren, 1987; Wong et al., 1989; He et al.,1991). The transport observed here is considerably less efficient than in

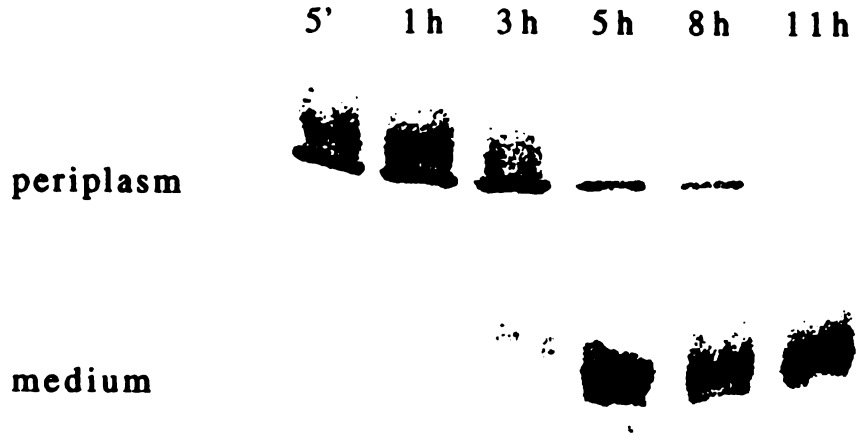


Figure 3.2. Pulse-chase of α-lytic protease in *E. coli*. Cultures were labeled for 5 minutes with ^{35}S-methionine and chased with cold methionine in fresh media. At the indicated times, aliquots were removed and fractionated (see Materials and Methods). Each fraction was immunoprecipitated with anti-α-lytic protease antibody on immobilized Protein A agarose beads. After SDS-PAGE, the bands were visualized by phosphorimaging.

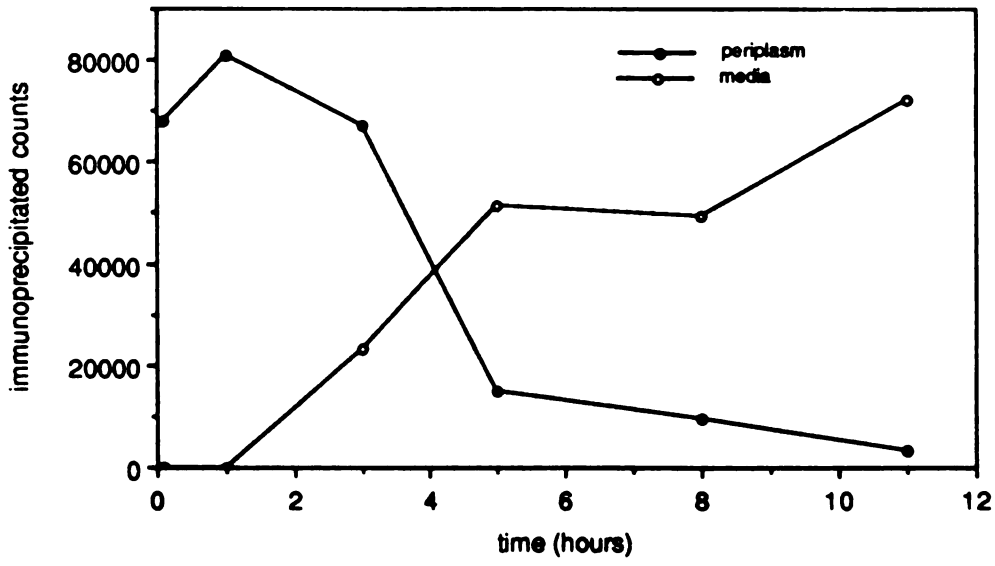


Figure 3.3. Quantitation of bands shown in figure 3.2. The dashed lines show the the halftime of export.

other Gram-negative bacteria. Nevertheless, when given a suitable substrate (such as α-lytic protease), *E. coli* is clearly capable of extracellular secretion via the two-step mechanism.

In the case of pullulanase secretion, fourteen native host (*Klebsiella pneumoniae*) proteins are required to reconstitute transport from the periplasm to the cell surface in *E. coli* (Possot et al., 1992). Since α -lytic protease accumulates (albeit slowly) in the media when heterologously expressed in *E. coli*, there does not appear to be a stringent requirement for accessory proteins specific to its native host. Nevertheless, we considered the possibility that such factors might exist to make the process more efficient in the native host than it appears to be in *E. coli*. Therefore pulse-chase experiments were carried out in *L. enzymogenes*.

Cell fractionation techniques for *Lysobacter* are not standardized, so we chose to analyze cell and media contents of α-lytic protease rather than periplasm and media content. Because *Lysobacter* is surprisingly resistant to standard cell lysis techniques, we utilized the export system itself to facilitate collection of the labelled cell contents. That is, after removal of the media fraction of each timepoint, the cells of that timepoint were allowed to secrete their cellular α-lytic protease content into fresh media. The data indicate that α-lytic protease is exported from its native host much more quickly than it is from *E. coli*, showing a halftime of ~3 minutes (Figures 4 and 5) .

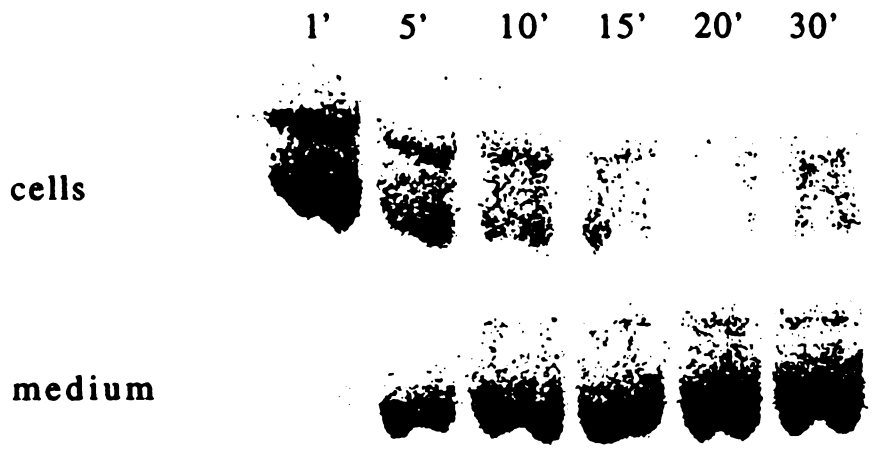


Figure 3.4. Pulse-chase of α-lytic protease in *L. enzymogenes*. Cultures were labeled for one minute and chased with cold methionine. At the indicated times, aliquots were removed and the cells and media were separated by centrifugation. Soluble cell contents were collected by a combination of export and lysis (see Materials and Methods). Proteins were immunoprecipitated with anti-α-lytic protease antibody on immobilized Protein A agarose beads. After SDS-PAGE, the bands were visualized by phosphorimager, shown above.

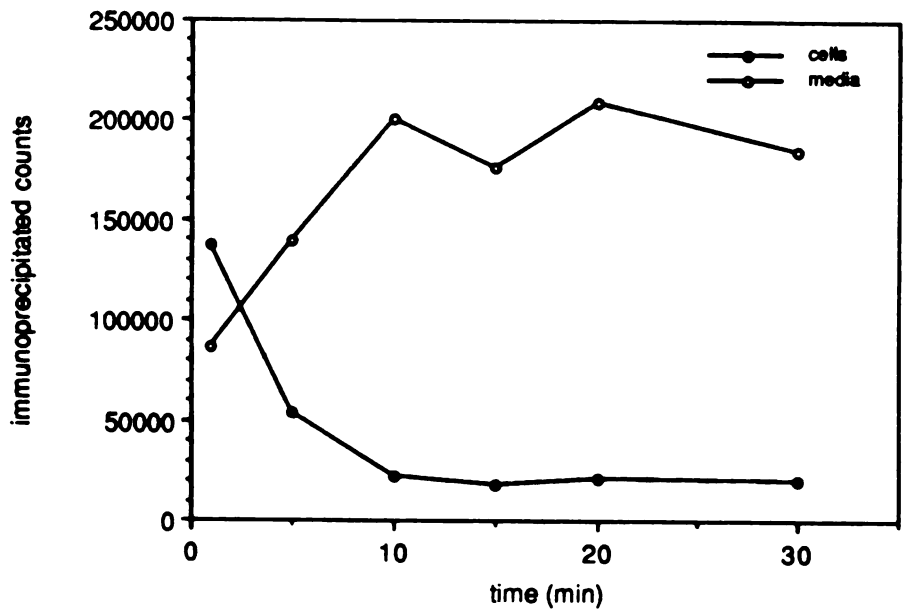


Figure 3.5. Quantitation of the bands shown in figure 4. The dashed lines show the the halftime of export.

Discussion

The data presented here indicate that α-lytic protease utilizes the two-step mechanism for export from *E. coli*. While the body of evidence is not large, it appears that those proteins bearing N-terminal signal sequences utilize a general secretion pathway that involves Sec proteins. Regardless of their final destination, these secreted proteins would be initially targeted to the periplasmic space. Certain secreted proteins lack a classical N-terminal signal sequence, and may have internal signal sequences; for these proteins, the one-step mechanism may be the preferred pathway (Mackman et al., 1985; Filloux et al., 1990; Wandersman and Delepelaire, 1990).

The halftime for extracellular secretion of α-lytic protease in *E. coli* is quite long, ~4 hours. This is considerably longer than has been observed for extracellular transport in the few other systems that have been characterized: The enzyme pullulanase, native to *Klebsiella pneumoniae*, is secreted from *E. coli* in less than five minutes (Pugsley, 1992); enterotoxin is secreted across the outer membrane of *Vibrio cholerae* with a halftime of ~13 minutes (Hirst and Holmgren, 1987); export of aerolysin from *Aeromonas salmonicida* takes between two and fifteen minutes (Wong et al., 1989; Wong and Buckley, 1989); pectate lyase is exported from *Erwinia chrysanthemi* in less than one minute (He et al., 1991).

There are several possible explanations for the slow export observed for α-lytic protease in *E. coli*. First, there is the fact that the expression system used for the pulse-chase experiments requires phosphate depletion of the cells. Arguing against this explanation is that the gross kinetics of export of α-lytic protease from plasmids with other promoters are similar to

that observed in this system (Agard lab, unpublished observations). Secondly, due to the temperature sensitivity of folding of α-lytic protease (Silen et al., 1989; Fujishige-Boggs et al., manuscript in preparation), these experiments were performed at 22°C. While the physical impact of lower temperature can only account for a factor of two or three in rates, it is possible that the biological impact is greater than this. Although the doubling time is indeed longer at 22°C, there are no obvious indications of detrimental changes in cell physiology (data not shown). A third possibility is that the native host, *L. enzymogenes*, provides specific accessory proteins to aid in transport across the outer membrane.

To test the latter hypothesis, the rate of export of α-lytic protease was examined in *L. enzymogenes*. The pulse chase data revealed that export from the native host is accomplished in a matter of minutes. While this may reflect differences in the media conditions mentioned above, precedent argues strongly for the presence of accessory proteins.

The heterologous expression of α-lytic protease in *E. coli* provides a facile system in which to examine factors influencing extracellular secretion. Notably, these factors are not absolutely required for export of α-lytic protease from *E. coli*. An exciting possibility is that there may be either a primitive or residual transport system in *E. coli* which α-lytic protease is exploiting for its secretion. The identification and overexpression of an *E. coli* extracellular transport system has obvious advantages. Therefore, we will use this system for the identification of extracellular secretion factors endogenous to *E. coli*, as well as for the characterization of accessory proteins from *L. enzymogenes*.

Acknowledgements

We are very grateful to L.L. Randall for training in pulse-chase methodology and for helpful discussions throughout this study, to Frank Loeffler for training in use of phosphorimager software, and to Hans de Cock for suggestions concerning labelling conditions under phosphate starvation. We are indebted to D. Littman and R. Grosschedl for the use of their phosphorimaging equipment and software. Financial support for this research was provided by the Howard Hughes Medical Institute. A.F. was supported by a graduate fellowship from the National Science Foundation.

References for Chapter Three: Kinetics of Export

Baker, D., Silen, J.L., and D.A. Agard. 1992a. A protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins* 12:339-344

Baker, D., Sohl, J.L., and D.A. Agard. 1992b. A protein-folding reaction under kinetic control. *Nature.* 365:263-265

Berkowitz, F.E. 1991. Bacterial toxins in the pathogenesis of infections. *Curr. Opinion in Infectious Disease.* 4:332-337

Birnboim, H.C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* 7:1513-1523

Chang, C.N., Kuang, W.-J., and E.Y. Chen. 1986. Nucleotide sequence of the alkaline phosphatase gene of *E. coli. Gene* 44:121-125

Deuschle, U., Kammerer, W., Gentz, R., and H. Bujard. 1986. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J.* 5:2987-2994

Eilers, M., and G. Schatz. 1986. Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature.* 322:228-232

Filloux, A., Bally, M., Ball, G., Akrim, M., Tommassen, J., and A.Lazdunski. 1990. Protein secretion in Gram-negative bacteria: transport across the outer membrane involves common mechanisms in different bacteria. *EMBO J.* 9:4323-4329

Fujishige, A., Smith, K.R., Silen, J.L., and D.A. Agard. 1992. Correct folding of α-lytic protease is required for its extracellular secretion from *Escherichia coli. J. Cell Biol.* 118:33-42

Galyov, E.E., Håkansson, S., Forsberg, Å., and H. Wolf-Watz. 1993. A secreted protein kinase of *Yersinia peudotuberculosis* is an indispensable virulence determinant. *Nature.* 361:730-732

Hall, M.N., Hereford, L., and I. Herskowitz. 1984. Targeting of *E.coli* beta-galactosidase to the nucleus in yeast. *Cell* 36:1057-1065

He, S.Y., Schoedel, C., Chatterjee, A.K., and A. Collmer. 1991. Extracellular secretion of pectate lyase by the *Erwinia chrysanthemi* Out pathway is dependent upon Sec-mediated export across the inner membrane. *J. Bacteriol.* 173:4310-4317

Hirst, T.R., and J. Holmgren. 1987. Transient entry of enterotoxin subunits into the periplasm occurs during their secretion from *Vibrio cholerae. J. Bacteriol.* 169:1037-1045

Horne, M. C. L. 1987. Thesis. San Francisco State University, San Francisco, CA

Krogfelt, K.A. 1991. Bacterial adhesin: genetics, biogenesis, and role in pathogenesis of fimbrial adhesins of *Escherichia coli. Revs. Infectious Diseases.* 13:721-735

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685

Mackman, N., Nicaud, J.M., Gray, L., and I.B.Holland. 1985. Identification of polypeptides required for the export of haemolysin 2001 from *E. coli.. Mol. Gen. Genet.* 201:529-536

Maniatis, T., Fritsch, E.F. , and J. Sambrook. 1982. *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory Press, N.Y.

Possot, O., d'Enfert, C., Reyss, I., and A.P.Pugsley. 1992. Pullulanase secretion in *Escherichia coli* K-12 requires a cytoplasmic protein and a putative polytopic cytoplasmic membrane protein. *Mol. Microbiol.* 6;95-105

Pugsley, A.P. 1992. Translocation of a folded protein across the outer membrane in *Escherichia coli. Proc. Nat. Acad. Sci. USA.* 89:12058-12062

Pugsley, A. P. 1993. The complete general secretory pathway in Gram-negative bacteria. *Microbiological Reviews.* 57:50-108

Randall, L.L., and S.J.S. Hardy. 1984. Export of protein in Bacteria. *Microbiological Reviews* 48:290-298

Randall, L. L., and S.J.S. Hardy. 1986. Correlation of competence for export with lack of tertiary structure of the mature species: a study *in vivo* of maltose-binding protein in *E. coli. Cell.* 46:921-928

Sadler, J., Tecklenburg, M.,and J. Betz. 1980 Plasmid containing many tandem copies of a synthetic lactose operator. *Gene* 8:279-300

Sen, K., and H. Nikaido. 1990. *In vitro* trimerization of OmpF porin secreted by spheroplasts of *Escherichia coli. Proc. Nat. Acad. Sci. USA.* 87:743-747

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., Frank, D., Fujishige, A., Bone, R., and D.A. Agard. 1989. Analysis of prepro-α-lytic protease expression in *Escherichia coli*

reveals that the pro region is required for activity. *J. Bact.* 171:1320-1325

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

Tommassen, J. and B. Lugtenberg. 1980. Outer Membrane Protein e of *Escherichia coli* K-12 is co-regulated with alkaline phosphatase. *J. Bact.* 143: 151-157

Wandersman, C., and P. Delepelaire. 1990. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc. Natl. Acad. Sci. USA* 87:4776-4780

Whitaker, D.R. 1970. The α -lytic protease of a myxobacterium. *In* Meth. Enz. 19:599-613

Wong, K.R. and J.T. Buckley. 1989. Proton motive force involved in protein transport across the outer membrane of *Aeromonas salmonicida*. *Science* 246:654-656

Wong, K.R., Green, M.J., and J.T. Buckley. 1989. Extracellular secretion of cloned aerolysin and phospholipase by *Aeromonas salmonicida*. *J. Bacteriol.* 171:2523-2527

Chapter 4: **Purification and in vitro Refolding of the Precursor**

Much of my early work in the Agard lab centered around purification of the precursor to α-lytic protease. As shown in Chapter Two, the precursor is not soluble but rather, is found in the outer membrane of E. coli, where it accumulates at temperatures (>30°C) that are nonpermissive for maturation.

The location of the precursor made purification difficult. We had first expected that the presence of a signal sequence would target the precursor in a soluble form to the periplasm. When standard osmotic shock treatment to release the contents of the periplasm gave very low and inconsistent yields of precursor, I reconsidered. First I tried gentle treatments, such as cell lysis followed by attempts to solubilize the precursor in nonionic detergents (0.1 and 1% Triton x-100, as well as Tween 20, β-octyl glucoside, and the mild ionic detergent deoxycholate, each at its critical micelle concentration. Next I tried very high and very low salts (0-1M NaCl in 200mM increments with 5 mM Phosphate buffer at pH 7), chaotropic salts, and variation of pH (4-10). To my great surprise, the precursor was largely insoluble with any of these treatments. The exceptions were deoxycholate and pH 10, each of which caused a *very* small percentage of the precursor to appear in the supernatant fraction. By this time, I began to believe what these results were telling me; the precursor was essentially insoluble, or inaccesssible.

In despair, I consulted Tack Kuntz, who gave me two pieces of advice. First he told me never to give up until I had tested the extremes, in this case a denaturing treatment. If that did not work, I would be able to give up with dignity, and peace of mind. If, however, it worked, I would be in much better shape to proceed. His reasoning was that it is always easier to work with a positive result, and he was absolutely right. His second piece of advice was to

consult Patricia Babbitt, then a senior graduate student whose thesis project had taken a path very similar to mine. Through Patsy, I obtained the modest literature then available on the subject of inclusion bodies. Certain proteins seemed for some reason to enter into dense, insoluble masses within the cell. These masses were dubbed inclusion bodies. The behavior of proteins in inclusion bodies matched what I had already noticed for the precursor. In much better spirits, I continued the purification effort.

Using a purification protocol for proteins that are located in insoluble inclusions within the cell helped to clarify the process. One's perspective takes a complete shift: in the purification of insoluble proteins, one can use that insolubilty as an advantage. Nevertheless, after washing away as many of the soluble proteins as possible, I noticed that the precursor did not behave as an inclusion body-protein would. Sonication caused partial release of the precursor into the aqueous fraction, but these populations could not truly be called soluble; they would not pass through 0.2 or 0.4 μm membranes, and they would not enter native gels.

I asked myself and members of the Walter lab what other insoluble material was present in the cell. Finally I hit upon the outer membrane. How had we missed this logical site? Most likely because it is called a membrane, and one expects protein of a membrane to be soluble in nonionic detergents. Not so with proteins of the outer membrane. A brief look at the literature revealed that isolation of these proteins required treatment with harshly denaturing detergents, sometimes for days or at elevated temperatures!

After that, the purification proceeded much more rapidly. I was able to combine the useful parts of inclusion body purifications and avoid the most harsh of treatments. In the final protocol, I purified the precursor by washing away the aqueous-soluble and nonionic detergent-soluble materials

first, and then solubilizing the remaining proteins under reversibly denaturing conditions. The final stages of purification were conducted completely under these denaturing conditions. Therefore, the following chapter represents the culmination of the original goal: purification and refolding of the precursor.

Together with my own work on the precursor, work on the pro region by lab members David Baker, Stephen Rader, and Julie Sohl gave us a new understanding of the temperature-sensitivity of folding (tsf) that had been observed in the expression of α-lytic protease in *E. coli.* In the following chapter we establish that the physical basis for the temperature-sensitivity of folding of α-lytic protease resides in the thermal instability of the pro region. To our knowledge, this is the first time that a physical basis has been assigned to a tsf phenotype. After further characterization, we intend to submit this work as a journal article.

Physical basis for temperature-sensitivity of folding of α-lytic protease.

Amy Fujishige, David Baker, and David A. Agard.

Departments of Biochemistry & Biophysics,

Pharmaceutical Chemistry and The Howard Hughes Medical Institute

University of California at San Francisco

San Francisco, CA 94143-0448

Abstract

α-Lytic protease is an extracellular serine protease produced by *Lysobacter enzymogenes* (Whitaker, 1970) that originates as a preproenzyme (Silen et al., 1988). It has been previously demonstrated that the large pro region is required both in vivo (Silen et al., 1989; Silen and Agard, 1989) and in vitro (Baker et al., 1992) for the proper folding of the protease. In addition, expression of α-lytic protease in *E. coli* has been shown to be temperature-sensitive in vivo (Silen et al., 1989), resulting in the accumulation of the proenzyme in the outer membrane at nonpermissive temperatures (Fujishige et al., 1992). The isolation and refolding of the intact wild type precursor to α-lytic protease is described. Refolding in vitro is accompanied by correct autoprocessing to the mature, active form. In addition, refolding in vitro displays temperature-sensitivity that parallels the in vivo phenotype. A similar temperature-sensitivity of refolding is seen when the pro region and the mature region are supplied as separate polypeptides. We present structural and functional evidence that, unlike the thermostable mature protease, the pro region is quite thermally labile. We propose that this is the physical basis for the temperature-sensitivity observed *in vivo.*

Introduction

Temperature-sensitivity of folding (tsf) has been observed for many proteins and protein variants. By definition, the tsf proteins and variants are those that acquire native structure only at permissive temperatures. Tsf proteins are distinct from temperature-sensitive (ts) proteins in that, once having acquired native structure, the tsf protein may be very thermostable. One model for the mechanism of the tsf phenotype invokes the presence of temperature-sensitive interactions along the folding pathway. In well characterized cases, it has been possible to associate the block in folding with a particular step in a folding pathway (Goldenberg et al.,1983). In cases characterized to date, this step has always preceded a (native) oligomerization. By far the most frequent result of an attempt to fold at the nonpermissive temperature is aggregation, or mis-oligomerization. Therefore, another model for the tsf phenotype invokes a temperature-sensitivity in the native monomer. In this model, nonpermissive temperatures would destabilize the monomer (possibly resulting in aggregation), while permissive temperatures would allow stable monomers to oligomerize, forming interactions that further stabilize the components. These two models are distinct in their implications, but are not mutually exclusive.

When expressed in *E. coli*, prepro-α-lytic protease displays temperature-sensitivity for folding: mature α-lytic protease (active enzyme that has been proteolytically processed to remove the both pre and pro regions) accumulates in the media at temperatures below 30°C, while inactive precursor accumulates in the cells at temperatures above 30°C (Silen et al., 1989). Inactivation of the protease by mutation of the active site serine to alanine (SA195) results in similar accumulation of cell associated precursor even at permissive

temperatures, indicating that α-lytic protease is self processing in *E. coli* (Silen et al., 1989). It has been established that the 166 amino acid pro region plays an obligatory role in the folding of the 198 amino acid protease domain: constructs lacking the pro region (Δ-pro-α-lytic protease) give rise to inactive α-lytic protease that is cell associated (Silen et al., 1989) whereas *in vivo* complementation of this Δ-pro molecule by co-expression of the pro region in trans gives rise to active α-lytic protease in the media (Silen and Agard, 1989). *In vitro* refolding of the denatured protease has only been possible on addition of the pro region (Baker et al., 1992a).

In the course of maturation, the precursor of α-lytic protease is proteolytically cleaved between the pro region and the protease region. These two parts form a tight heterodimeric complex ($K_D \sim 10^{-10}$M; Baker et al, 1992a). The bound pro region inhibits activity of the protease. Further degradation of the pro region results in its dissociation from the protease region, leaving the monomeric, active protease. Mature α-lytic protease is very thermostable, but is clearly sensitive to temperature during folding *in vivo* and *in vitro.*

We report here the purification and in vitro refolding of pro-α-lytic protease. Refolding is temperature sensitive in vitro, with a similar transition temperature as is seen in vivo. Furthermore, one component of the heterodimer, the pro region, displays thermal instability that could account for the observed tsf phenotype.

Materials and Methods

Media, Strains, and Plasmids-- plasmid pALP5 contains a copy of the gene for pro-α-lytic protease fused in frame to the alkaline phosphatase signal sequence. This construct is driven by the PhoA promoter and has been described previously (Silen et al., 1989). pALP12 contains the same pre-pro-α-lytic protease construct driven by the IPTG-inducible P_{A1} early T7 promoter (Deuschle et al., 1986). pGEXPRO4 is an expression vector for a fusion protein with glutathione-S-transferase at the N-terminus and the pro region of α-lytic protease at the C-terminus. The final two amino acids of the pro region were changed in cloning from Thr-Thr to Lys-Leu-Lys. The construction is in other ways similar to pGEXPRO, which was described previously (Baker et al., 1992a).

Protein Puification– Mature α-lytic protease was purified as described previously (Hunkapiller et al., 1973; Baker et al., 1992a). Precursor was isolated from *E. coli* MH1 cells (*araD139 ΔlacX74 galU galK hsr rpsL*; M.Hall, 1984) harboring plasmid pALP5 (Silen et al., 1989) and grown at 37°C. The purification process was as follows: 25 g cell paste was resuspended in 50 ml STET buffer (8% sucrose, 45 mM EDTA, 5% TritonX-100, 50 mM Tris, pH 8; ref, date) with 0.5 g lysozyme (10 mg/ml final concentration). The cells were stirred and allowed to lyse at room temperature. DNAse was added to a final concentration of 0.04 mg/ml with 1 ml 1M $MgCl_2$ (20 mM final concentration) to reduce viscosity. The mixture was centrifuged at 17,000xg for 20 minutes at 4°C. The supernatant was discarded and the pellet was retained. The pellet was washed in 2% β-octyl glucoside, 5 mM KPO_4, pH 7 at 37°C for one hour, and rinsed once in 5 mM KPO_4, pH 7. The washed pellet was then combined with 2.4 g solid crystalline urea and brought to a total volume of 5 ml with 5mM KPO_4, pH 7, 100 mM DTT (final concentration of urea equal to 8M). Unsolubilized

materials were removed by centrifugation (20,000 xg, 40 min at 15°C), and 800 µl of the supernatant was loaded in onto a BioRad MA7C HPLC ion exchange column that had been preequilibrated in 8M urea, 0.1 mM DTT, 5 mM KPO_4, pH 7 (Buffer A). After washing in Buffer A, the protein was eluted in a 0-30% gradient with Buffer B (Buffer A + 0.5 M sodium acetate). The fractions were characterized by SDS-PAGE, immunoblot, and α-lytic protease activity upon refolding (see below). Pooled fractions from the MA7C column were exchanged into guanidine HCl using a Centricon™ 10 (Amicon, Division of W.R. Grace and Co., Danvers, MA) and applied to a Superose 12 HPLC gel filtration column (Pharmacia, Uppsala, Sweden).

pGEXPRO4-bearing cells were lysed and the lysate applied to glutathione agarose (Smith and Johnson, 1988) for initial purification of the fusion protein. The protein was then cleaved with FactorX and applied at pH 10 to a Poros (PerSeptive Biosystems, Cambridge, MA) ion exchange column. The pro region was eluted at ~50mM NaOAc in a 0-100 mM gradient. Pooled fractions were Centriconed to concentrate the pro region and to exchange the buffer to $NaPO_4$, pH 8.

Gel Electrophoresis of Proteins and Immunoblot Analysis-- SDS-PAGE was carried out on a Hoefer minigel apparatus. Samples were combined with Laemmli sample buffer (Laemmli, 1970), boiled for one minute, and microfuged for one minute. The supernatant was run on a discontinuous (5%/12%) SDS-polyacrylamide gel. Where appropriate, the contents of SDS polyacrylamide gels were transferred to nitrocellulose using an ABN polyblot apparatus. Immunoblots were probed with affinity-purified rabbit anti-α-lytic protease antibody or an antibody directed against a fusion protein of the pro region with glutathione transferase (Baker et al., 1992a) followed by

alkaline phosphatase conjugated-goat-anti-rabbit antibody (Bio-Rad Laboratories, Richmond, CA).

Refolding: Effect of pH, Ionic strength, Redox, Additives and Temperature– Based on preliminary studies, refolding was carried out by dilution at room temperature. Typically 10 μl of the purified, denatured precursor was diluted 100-fold to approximately micromolar concentration into the indicated buffer with 1 mM substrate (succinyl-alanyl-prolyl-alanyl-para-nitroanilide; SAPApNA) present. The refolding was terminated by the addition of trypsin (final concentration10μm/ml), which destroys the pro region but leaves the folded protease intact; trypsin does not interfere with the α-lytic protease assay (Baker et al., 1992a). Recovered activity was measured as the absorbance at 410nm from released paranitroaniline. Substrate was present throughout the refolding process.

The amount of time before trypsinization was varied from several minutes to twelve hours.

The pH dependence of refolding was assessed by dilution as described above into 100mM buffers at pH 4, 5, 6, 7, 8, and 9. Sodium acetate was the buffer at pH 5 and 6, potassium phosphate was used at pH 7, and Tris-HCl was used at pH 8 and 9.

The effect of ionic strength on folding was assessed by dilution into potassium phosphate at 0, 5, 50, 100, 200, or 500 mM concentration, in the presence of 1 mM substrate (succinyl-alanyl-prolyl-alanyl-para-nitroanilide; SAPApNA). The refolding was terminated by the addition of trypsin after 15 minutes.

Redox conditions were varied using 1-100mM glutathione (oxidized:reduced = 1:10). Various proteins were added in an attempt to stabilize the nascent folded protein. In separate experiments, 1 mg/ml bovine serum

albumin, DnaK, GroEL+/- GroES, Hsp70 and GrpE were included in the refolding buffer, with or without 1 mM ATP.

The effect of temperature on refolding was assessed by carrying out the optimized refolding at various temperatures. The solutions were then cooled or heated to 25°C for the activity assay.

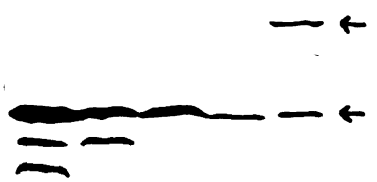
Circular Dichoism of the Pro Region-- Purified pro region was brought to 0.1 mg/ml in 20mM $NaPO_4$, pH 8 buffer. Circular dichroism was measured using a Jasco J500A spectrophotometer using a 0.1mm pathlength. Initial scans indicated a high α-helical content. The ellipticity was then measured at 220 nm as a function of temperature.

Results

α-Lytic protease is native to *Lysobacter enzymogenes*, a soil bacterium whose maximal temperature of growth is ~30° C. Heterologous expression of the protease in *E. coli* is successful, but is temperature-sensitive. The transition temperature appears to be around 30° C. Expression, processing, secretion and activity are normal below this temperature and are attenuated above this temperature (Figure 4.1; see also Silen et al., 1989; Fujishige et al., 1992). The goal of this study was to determine whether this observed phenotype could be accounted for physicochemically. Although it was possible that an extragenic (e.g. heat shock) factor played a role, the coincidence of transition temperature and maximal growth temperature of the native host suggested that the observed temperature-sensitivity was intrinsic to this protein. Therefore our initial studies focussed on the folding process.

Folding of α-lytic protease is pro region dependent. It has not been possible to denature and refold the protease in the absence of the pro region. We have approached the refolding with two systems. In a previous approach, we isolated pro region and combined it with denatured, purified mature α-lytic protease to effect pro region dependent folding in complementation(Baker et al., 1992a, b). In another approach, reported below, we have isolated and refolded the uncut proenzyme from cells that are grown at nonpermissive temperatures. In this study, we examine the temperature-dependence of refolding for each of these systems.

Purification of the wild type precursor to α-lytic protease from the outer membrane of *E. coli* was accomplished in three major steps under denaturing conditions. First, cells were lysed, and most water- and nonionic detergent-soluble materials were washed away. Next, a urea-DTT soluble



Growth Temperature (°C): 15 18 22 28 32

Figure 4.1 Temperature-dependence of in vivo expression of α-lytic protease in *E. coli*. pALP12 cells were grown at the indicated temperatures. The cell contents were applied to an SDS gel and immunoblotted. The arrows indicate the precursor and mature forms. Only mature forms showed α-lytic protease activity; those containing any amount of precursor were inactive.

fraction was chromatographed under denaturing conditions on an ion exchange column. Finally, the pooled fractions of this column were applied to a gel filtration column, again under denaturing conditions. We have noted that both native and denatured mature α-lytic protease have a tendency to stick to gel filtration columns. Indeed, in the final step of purification of the precursor, we were obliged to change the running/denaturing buffer from urea to guanidine in order to minimize sticking. The resulting protein is >95% pure.

The fractions from each step were assayed by immunoblot and refolding (Figure 4.2a, b). Only the fractions that contained immuno-cross reactive protein species showed α-lytic activity upon refolding. The estimated recovery from the refolding step is 1%. While this low yield precludes physical measurements of the refolding process, recovered protein is still easily quantitated by our activity assay. Refolding was terminated after 15 minutes (Fig. 4.3) by the addition of trypsin, which destroys the pro region and any unfolded molecules, but leaves mature α-lytic protease intact.

Slight improvements were achieved by adjusting the composition of the buffer. Varying the redox conditions did very little to improve either the rate or the yield of refolding. This is surprising, in light of the three disulfide bridges present in mature molecule. Under improved refolding conditions, it is expected that redox conditions will have a greater effect. Neither alteration of ionic strength nor the addition of chaperones increased the refolding yield (data not shown). The most notable effects came with variation of pH (Fig 4.4) and temperature (Fig 4.5) during refolding. pH 7 appears to be optimal for refolding of the precursor.

The temperature-dependence of refolding the precursor parallels what is seen *in vivo*. Refolding in complementation (pro+denatured mature) was

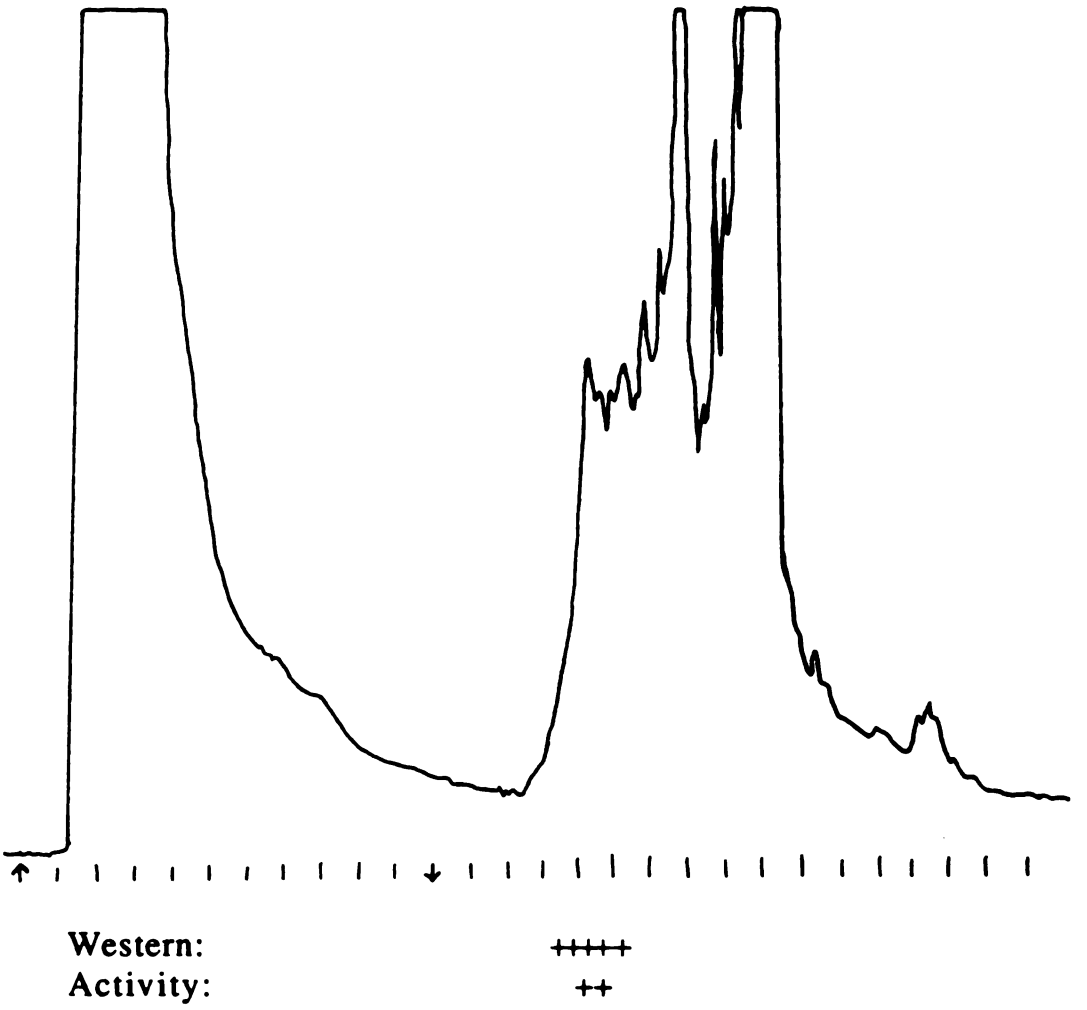


Figure 4.2a. Purification and Refolding of the 37°C precursor of α-lytic protease. The total eluted protein from the MA7C columnwas monitored by absorbance at 280 nm (shown on trace). Fractions were assayed for anti-α-lytic protease immuno-cross reactive protein by western blot and were assayed for α-lytic protease activity upon refolding (positive signal in assay indicated under fractions by plus signs). The active site mutant, SA195, was used as a negative control. Upward arrow indicates time of injection; downward arrow indicates beginning of gradient and fraction collection. Two fractions were collected per marked time interval.

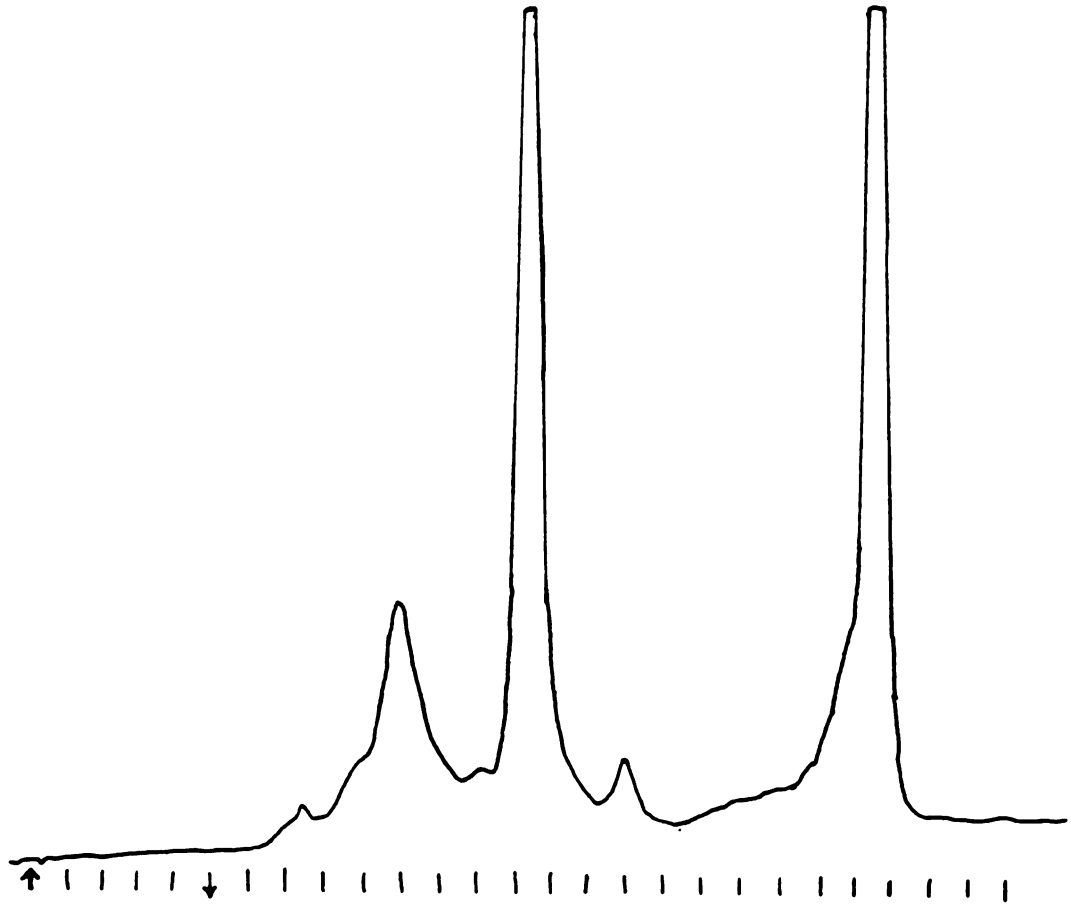


Figure 4.2b. Purification and Refolding of the 37°C precursor of α-lytic protease. The total eluted protein from the Superose 12 column was monitored by absorbance at 280 nm (shown on trace). Fractions were assayed for anti-α-lytic protease immuno-cross reactive protein by western blot and were assayed for α-lytic protease activity upon refolding (positive signal in assay indicated under fractions by plus signs). Upward arrow indicates time of injection; downward arrow indicates beginning of fraction collection. Two fractions were collected per marked time interval.

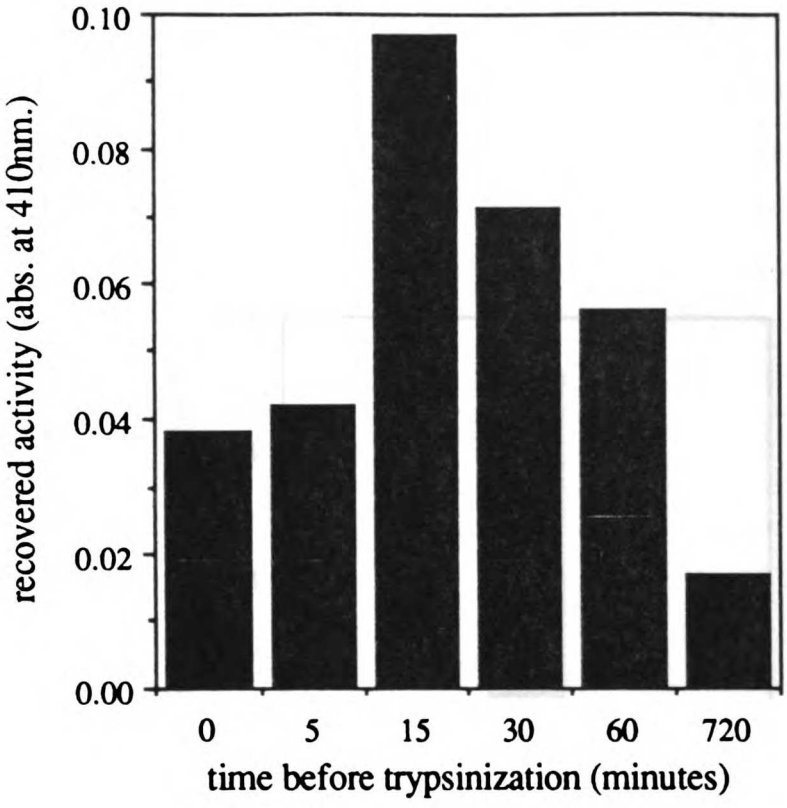


Figure 4.3. Trypsinization timing optimum. Purified, denatured precursor was allowed to refold for the indicated times before being trypsinized. Trypsin degrades the pro region as well as unfolded molecules, but does not affect folded α-lytic protease. Refolded samples were assayed for α-lytic protease activity after trypsinization.

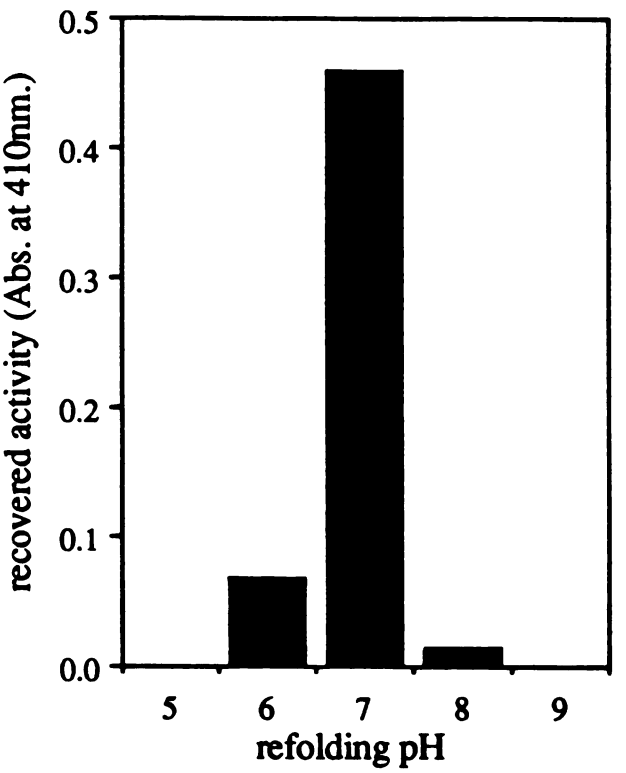


Figure 4.4. Refolding: pH Optimum. The precursor was allowed to refold at the indicated pH (see text), after which a portion was assayed in standard assay buffer conditions for α-lytic protease activity.

similarly temperature-sensitive (Fig. 4.5), indicating that proteolytic processing between the pro region and the protease region is not the temperature-sensitive step. Since mature α-lytic protease is quite thermostable, we examined the thermostability of the isolated pro region. In marked contrast to the mature protease, the pro region rapidly loses secondary structure (as measured by circular dichroism) with increasing temperature (Fig 4.6).

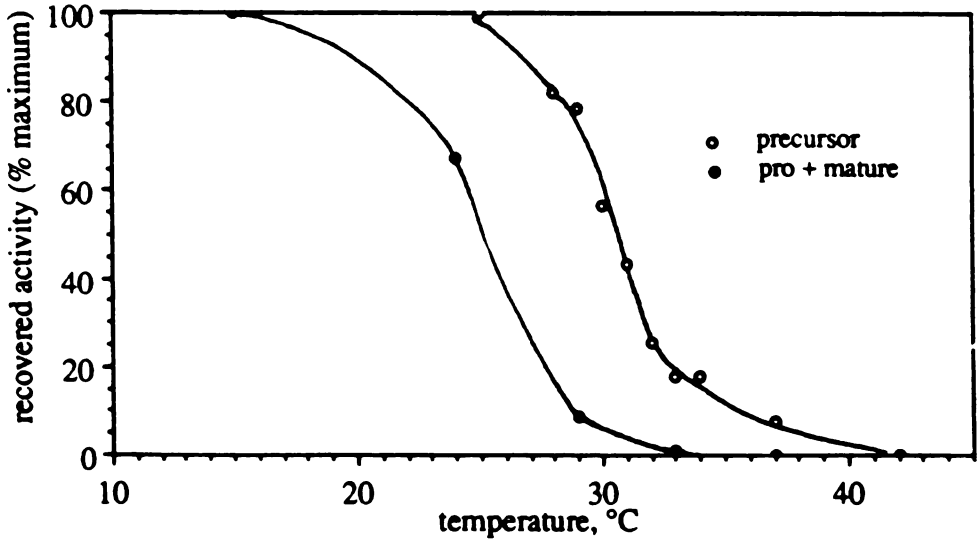


Figure 4.5. Temperature-sensitivity of folding of the full length precursor of α-lytic protease and the complementation system (pro+mature). The precursor or pro+mature were allowed to refold at the indicated temperatures (see text). After trypsinization a portion was assayed in standard assay buffer and temperature conditions for α-lytic protease activity.

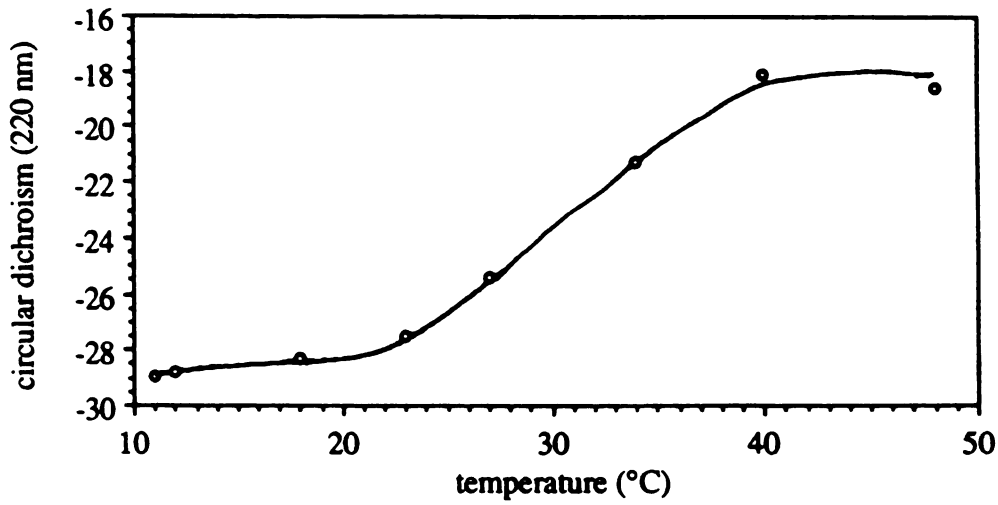


Figure 4.6. Thermal melt of the pro region measured by circular dichroism. The ellipiticity of the pro region at 220nm was measured at the indicated temperatures.

Discussion

In this report we have described the purification and refolding of a precursor of α-lytic protease. Our initial characterization of the refolding revealed marked pH and temperature sensitivities. Because the temperature-sensitivity parallels that seen in vivo, we pursued this aspect of folding. We found that the temperature sensitive component in the folding of α-lytic protease is the pro region. This result is consistent with the observed tsf phenotype as well as the thermostability of the mature protease.

Unprocessed and misfolded forms of α-lytic protease are located in the outer membrane of *E. coli* and it was therefore not possible to solubilize the precursor in nondenaturing buffers (Fujishige et al., 1992). Purification of outer membrane proteins typically involves prolonged exposure to SDS or other harsh denaturants. In the method described here, the outer membrane-associated precursor was solubilized in urea rather than ionic detergent in order to facilitate refolding. Refolding is accompanied by autoprocessing (data not shown), as would be predicted by previous results from experiments with the active site mutant, SA195 (Silen et al, 1989).

Despite the presence of three disulfide bonds in the mature protease, variation of redox conditions did not improve the refolding yield. At present the refolding yield is quite low for the fully reduced precursor, but is better when denatured (but unreduced) mature protease is combined with purified pro region (Baker et al., 1992a). Presumably, there are early folding steps (preceding disulfide bond formation) that have not yet been optimized in the refolding of the full length precursor. Given this, it is not surprising that the pH optimum for refolding occurs at pH 7 (rather than pH 8 or higher, where disulfide bond formation is more favorable). When the requirements for the

early steps have been met, one would expect that the redox conditions will become important determinants of refolding rate and/or yield.

The marked temperature-sensitivity of folding observed in vivo is clearly paralleled by temperature-sensitivity of folding in vitro. Our data indicate that the temperature-sensitivity of folding seen in α-lytic protease is not an artifact of heterologous expression in *E. coli*, but is rather a physical property inherent to the precursor.

The two component in vitro system (pro+mature) also shows temperature-sensitivity, with a slightly lower permissive temperature than for the full length precursor or the in vivo system. While in the full length precursor the pro region has a stoichiometric local concentration with respect to the protease region, the two component system requires the two polypeptides to encounter each other as well as to bind to each other. The discrepancy in temperature-sensitivities may indicate that in the two component system, the pro region has a better chance of noncovalent interaction with the mature region at lower temperatures where more structure is present.

The mature protease is extremely thermostable. Interestingly, our latest data indicate that the pro region loses secondary structure at ~ 30°C. This melting temperature coincides with the transition temperatures for permissive folding both in vivo and in vitro. Indeed, this is the maximal growth temperature of the native host of α-lytic protease, suggesting that the folding has been optimized to the growth range of the host.

Hitherto, proteins with tsf phenotypes have been characterized temporally; that is, the temperature-sensitivity has been assigned to a particular step in the folding/assembly pathway. We have determined the temperature-sensitive component in a spatial rather than a temporal sense:

the thermal instability of the pro region is entirely sufficient to explain the tsf phenotype. Although α-lytic protease is a monomer in its mature form, the maturation pathway includes the formation of a heterodimer (pro + mature). One component of this heterodimer, the pro region, has a true ts phenotype. Since the pro region is only transiently required during folding but is not present in the mature molecule, this ts phenotype is observed as a tsf phenotype of the other component of the heterodimer, α-lytic protease. Thus, we have an example of the unstable monomer model for temperature-sensitivity of folding. While much more difficult to detect, the theoretical presence of temperature-sensitive, nonnative folding transitions is uncompromised by these results.

Acknowledgements

We are grateful to Stephen Rader and Julie L. Sohl for their contributions in obtaining circular dichroism measurements of the pro region, for their supporting data, and for helpful discussions throughout this work. We thank the Dept. of Pharmaceutical Chemistry at U.C.S.F. for the use of the Jasco Spectrophotometer. Special Thanks to A. Frankel and his laboratory at The Gladstone Institute, San Francisco, CA, for their instruction and the use of their Aviv Spectrophotometer.

References for Chapter 4: Temperature-Sensitivity

Baker, D., Silen, J.L., and D.A. Agard. 1992a. A protease pro region required for folding is a potent inhibitor of the mature enzyme.*Proteins* 12:339-344

Deuschle, U., Kammerer, W., Gentz, R., and H. Bujard. 1986. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J.* 5:2987-2994

Fujishige, A., Smith, K.R., Silen, J.L., and D.A. Agard. 1992. Correct folding of α-lytic protease is required for its extracellular secretion from *Escherichia coli. J. Cell Biol.* 118:33-42

Goldenberg, D.P., Smith, D.H., and J. King. 1983. Genetic analysis of the folding pathway for the tail spike protein of phage P22. *P.N.A.S. (U.S.A.)* 80:7060-7064

Hall, M.N., Hereford, L., and I. Herskowitz. 1984. Targeting of *E.coli* beta-galactosidase to the nucleus in yeast. *Cell* 36:1057-1065

Hunkapiller, M. W., S. H. Smallcombe, D. R. Whitaker, and J. H. Richards. 1973. Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implications for the catalytic mechanism of serine proteases. *Biochemistry* 12:4732-4743.

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., Frank, D., Fujishige, A., Bone, R., and D.A. Agard. 1989. Analysis of prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro region is required for

activity. *J. Bact.* 171:1320-1325

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

Smith, D.B. and K.S. Johnson. 1988. Single-step purification of polypeptides espressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* 67:31-40

Whitaker, D. R. 1970. The α-lytic protease of a Myxobacterium. *Methods in Enzymol.* 19:599-613.

Chapter 5: **Structure Function Studies**

The relationship between structure and function of proteins has been pursued on many levels and with many techniques and systems over the years. In the Agard lab, we have approached this problem by examining the structural parameters that determine substrate specificity for α-lytic protease.

The primary specificity of α-lytic protease is for small hydrophobic residues at the P1 position. The S1 site of the binding pocket of α-lytic protease is lined with the side chains of Met 192, Met 213, and Val 217A (chymotrypsin numbering). Using site-directed mutagenesis, these residues have been changed singly and in combination. Initially, each of the methionines were changed to alanine. The removal of a mass equivalent in volume to three methylenes from the S1 binding site was expected to create the same amount of space for a larger P1 sidechain; that is, the combination of methionine in the enzyme and alanine in the substrate was expected to be equivalent to alanine in the enzyme and methionine in the substrate.

For the single mutant, Met 192→ Ala (MA192), this proved basically to be the case. In a set of enzyme kinetics assays on a set of tetrapeptide substrates with an array of amino acids in the P1 position, the mutant enzyme showed high activity for substrates with methionine in the P1 position. In fact, the MA192 mutant showed even greater activity towards tetrapeptide substrates with phenylalanine in the P1 position. This indicated that the already enlarged binding pocket was able to utilize energetically allowed conformational changes to expand even beyond what had been predicted from the static picture. High resolution crystal structures of the mutant enzyme alone and in complex with any of several transition state analogs have confirmed this. Curiously, the enlarged pocket of MA192 was also able to

accommodate small hydrophobics in the P1 position without considerable loss of activity. This indicated that collapse of the binding pocket to bind a smaller substrate was also energetically allowed. Again, this was confirmed by x-ray crystallography.

Although its enzymatic activity was somewhat lower than that of MA192, the other methionine mutant (Met 213→ Ala: MA213) showed a similarly broad specificity. I endeavored to expand our understanding of these highly active, broadly specific enzymes by solving the structure of MA213 when challenged with the largest (phenyalanine) or the smallest (alanine) natural sidechain at the P1 position of a tetrapeptide boronic acid inhibitor, and comparing this structure to the enzyme kinetics obtained for the analogous tetrapeptide p-nitroanalide substrate.

A new understanding of protein-ligand interactions has emerged from the data obtained with these two binding pocket mutants. Our current view is that there are areas of structural plasticity in the wild type enzyme that are exploited to a higher degree in the mutant enzymes. This plasticity allows the enzyme to utilize hydrophobic interaction, hydrogen bonding, van der Waals interaction and other physical phenomena in order to bind ligands of different size or character.

The MA213 structure and function studies mentioned above were published in the context of a larger body of work (Biochemistry, 1991, Volume 30, pages 10388-10398). I have included this paper in its entirety in the following pages.

Structural Basis for Broad Specificity in α-Lytic Protease Mutants [1]

Roger Bone [1 ||], Amy Fujishige [1], Charles A. Kettner [§], David A. Agard [* 1]

Departments of Biochemistry & Biophysics, Pharmaceutical Chemistry, and The Howard Hughes Medical Institute
University of California at San Francisco
San Francisco, CA 94143-0448

and

Du Pont Merck Pharmaceutical Company
Du Pont Experimental Station
Wilmington, Delaware 19880-0328

Running title: Structural basis for Broad Specificity

FOOTNOTES

[1] This work was supported by funds from the Howard Hughes Medical Institute and from an NSF Presidential Young Investigator grant (DAA). R. Bone was supported by NIH National Research Service Award GM11174-02.

[1] University of California, San Francisco

[§] Du Pont Merck Pharmaceutical Company

[||] Current Address: Department of Biophysical Chemistry, Merck Sharp & Dohme Research Laboratories, Box 2000 (RY80M-136), Rahway, NJ 07065

[*] Author to whom correspondence should be addressed.

1. Abbreviations: sAAP-, succinyl-Ala-Ala-Pro-; mAAP-, *methoxysuccinyl-Ala-Ala-Pro-, bAP, *tert*-butyloxycarbonyl-Ala-Pro-; MeOSuc-, methoxysuccinyl-; Tris, tris(hydroxymethyl)-aminomethane; *p*-Na, *para*-nitroaniline; Norleu, norleucine; the prefix "boro" of -boroVal- indicates that the carbonyl of the amino acid residue, in this case a valyl residue, is replaced by $-B(OH)_2$; The systematic name for valine boronic acid would be (1-amino-2-methylpropyl)boronic acid; RMS, root mean square.

2. In protease substrate nomenclature the residue to the N-terminal side of the scissile bond is the P_1 residue, the next residue toward the N-terminus is the P_2 residue, etc. I. Schecter and A. Berger, *Biochem. Biophys. Res. Commun. 27*, 157 (1967).

3. Residues in α-lytic protease are numbered by homology with chymotrypsin (Fujinaga *et al.*, 1985) and range from 15A-244.

4. Boron bond angles of 109° would be expected in a purely tetrahedral adduct, while angles of 120° would be expected in a trigonal planer adduct.

5. The energy of hydrogen bond formation also decreases if hydrogen bonds

become shorter than optimal. In this series of enzyme-inhibitor complexes one hydrogen bond (Gly 193 N--P_1 O_1) improves by lengthening to a more optimal distance as the size of the P_1 side chain increases. However, this is compensated by a corresponding decrease in the length of another hydrogen bond (His 57 $N\varepsilon_2$--P_1 O_2) to values shorter than optimum.

6. Hydrogen bonds stretched the most include not only hydrogen bonds in the oxyanion hole, but also hydrogen bonds involving residues 214 and 216. The release of strain induced by bad steric contacts is complex, but appears to involve pivoting of the peripheral portion of the inhibitor about the P_1 boron. This may be due in part to the packing of P_1 side chains against the P_2 Pro of the inhibitor. In a recent comparison of the binding of transition state analogs and product analogs, it was shown that the only hydrogen bond disrupted in the product complex was the hydrogen bond with residue 214 (Bone *et al.*, 1991c).

Abstract

Binding pocket mutants of α-lytic protease (Met 192 -> Ala and Met 213 -> Ala) have been constructed recently in an effort to create a protease specific for Met just prior to the scissile bond. Instead, mutation resulted in proteases with extraordinarily broad specificity profiles and high activity [Bone, R., Silen, J. L., & Agard, D. A. (1989) *Nature 339*, 191-195]. To understand the structural basis for the unexpected specificity profiles of these mutants, high resolution x-ray crystal structures have been determined for complexes of each mutant with a series of systematically varying peptidyl boronic acids. These inhibitory analogs of high energy reaction intermediates provide models for how substrates with different side chains interact with the enzyme during the transition state. Fifteen structures have been analyzed qualitatively and quantitatively with respect to enzyme-inhibitor hydrogen bond lengths, buried hydrophobic surface area, unfilled cavity volume, and the magnitude of inhibitor accommodating conformational adjustments (particularly in the region of another binding pocket residue, Val 217A). Comparison of these four parameters with the K_i of each inhibitor and the k_{cat} and K_m of the analogous substrates indicates that while no single structural parameter consistently correlates with activity or inhibition, the observed data can be understood as a combination of effects. Furthermore, the relative contributions of each term differs for the three enzymes, reflecting the altered conformational energetics of each mutant. From the extensive

structural analysis, it is clear that enzyme flexibility, especially in the region of Val 217A, is primarily responsible for the exceptionally broad specificity observed in either mutant. Taken together, the observed patterns of substrate specificity can be understood to arise directly from interactions between the substrate and the residues lining the specificity pocket, and indirectly from interactions between peripheral regions of the protein and the active site region that serve to modulate active-site flexibility.

Introduction

One of the fundamental functions of an enzyme is to provide specificity by limiting the range of substrates which are catalytically productive. Since catalysis is the result of an enzyme's ability to bind the transition state for a reaction (Wolfenden, 1972; Fersht, 1985a), specificity must derive from the selective binding of the transition states for the reactions of favored substrates. In empirical terms, the selection of substrates must result from the balance between energetically favorable intermolecular interactions (hydrogen bonds, solvent exclusion from hydrophobic surfaces, electrostatics, van der Waals interactions) and energetically costly consequences of accommodating substrates in an imprecisely tailored active site (hydrogen bond distortion or loss, cavity formation, induced conformational changes). Functional groups participating in these interactions, and estimates for their energetic contributions, can be identified by studying how catalytic efficiency is influenced by the systematic variation of substrate and enzyme structure (Fersht, 1988; Fersht, 1985b; Estell *et al.*, 1986). However, a thorough understanding of how the favorable and unfavorable interactions of important structural elements, contributed by both the enzyme and substrate, strike a balance to produce a particular pattern of specificity can only be achieved through direct structural analysis.

Recently the specificity of α-lytic protease, an extracellular serine protease of *L. enzymogenes* (Whitaker, 1970), has been examined using a combination of crystallographic and mutational analyses. On small peptide substrates, sAAP-X-*p*-Na[1] where sAAP is succinyl-Ala-Ala-Pro- and X is Ala, Val, Met, Leu, or Phe, the primary specificity of the enzyme is for Ala in the P_1[2] position with activity falling off dramatically as the size of the P_1

substrate side chain increases (Table 5-I; Bone *et al.* 1989a; Bauer *et al.*, 1981). Crystallographic analysis of a series of complexes between wild type α-lytic protease and inhibitory peptide boronic acids, which mimic the transition state for substrate hydrolysis, revealed why substrates with P_1 side chains increasing in size are selected against (Fig 5.1; Bone *et al.*, 1989b). As the size of the P_1 side chain increases, hydrogen bonds become distorted and conformational shifts of greater magnitude are required for accommodation of the P_1 side chain. Since there is no compensating increase in the amount of hydrophobic surface area buried upon inhibitor interaction with the enzyme, binding of inhibitors with large P_1 side chains is unfavorable. It appeared that specificity was restricted to substrates with Ala in the P_1 position because of the presence in the primary specificity pocket of two Met residues (Met 192 and Met 213[3]) that block what would otherwise be a large hydrophobic pocket.

To test this hypothesis, each of the two Met residues was mutated to Ala (Bone *et al.*, 1989a). It was reasoned that the resulting mutants would have specificity for substrates with Met in the P_1 position which would complement the Met -> Ala mutations. Remarkably, the mutagenesis resulted in two proteases having extraordinarily broad specificity, one of which retained very high activity. In both cases, activity did not decrease as expected for substrates with small side chains in the P1 position. In addition, surprising increases in activity were observed for substrates with the largest and bulkiest P_1 residues (Leu, Phe). Initial structural analyses of these mutants suggested that one of these enzymes (Met 192 -> Ala) was able to utilize conformational changes in a productive manner to accommodate a substrate with Phe in the P_1 position (Bone *et al.*, 1989a). Furthermore, the structures of the free mutants revealed that the enzymes had unexpectedly small hydrophobic pockets.

Table 5-I: Values of Kinetic Parameters for Wild-Type and Mutant α-Lytic Proteases[a]

P_1 Residue	Wild Type k_{cat} (s^{-1})	Wild Type K_M (mM)	Wild Type k_{cat}/K_M ($s^{-1}M^{-1}$)	Wild Type K_i (nM)	MA192 k_{cat} (s^{-1})	MA192 K_M (mM)	MA192 k_{cat}/K_M ($s^{-1}M^{-1}$)	MA192 K_i (nM)	MA213 k_{cat} (s^{-1})	MA213 K_M (mM)	MA213 k_{cat}/K_M ($s^{-1}M^{-1}$)	MA213 K_i (nM)
Gly	12 (2)	35 (10)	330 (25)	---	0.46 (0.03)	15 (1)	32 (1.3)	---	3.1 (0.4)	40 (5)	78 (1.3)	---
Ala	75[b]	3.6[b]	21000[b]	67[d]	37[c]	3.6[c]	10000[c]	64 (2)	34[c]	57[c]	600[c]	270 (3)
Val	13[b]	16[b]	790[b]	6.4[d]	3.4[c]	1.1[c]	3000[c]	1.3[e] (0.1)	10[c]	29[c]	340[c]	210[c]
NorLeu	3.7 (0.07)	49 (10)	75 (2)	1100[b]	140 (3)	0.14 (0.01)	970000 (40000)	0.26[e] (0.01)	26 (3)	16 (2)	1600 (30)	50 (1.3)
Met	56[b]	31[b]	1800[b]	---	120[c]	0.33[c]	350000[c]	---	---	---	980[c]	---
Leu	1.2[b]	290[b]	4.1[b]	2000 (20)	87[c]	0.77[c]	110000[c]	0.58[e] (0.03)	---	---	160[c]	66 (1)
Phe	0.0068[b]	17[b]	0.38[b]	540[d,f]	130[c]	0.40[c]	310000[c]	0.60[c]	47[c]	16[c]	340[c]	240 (4)

[a] The substrates used were sAAP-X-*p*-Na, where sAAP- is succinyl-Ala-Ala-Pro-, X is Gly, Ala, Val, Norleu, Met, Leu or Phe and *p*-Na is *p*-nitroaniline. The inhibitors used were mAAP-boroX, where mAAP- is methoxysuccinyl-Ala-Ala-Pro- and boroX is the α-amino boronic acid analog of Ala, Val, Norleu, Leu or Phe. For values determined in this work standard deviations (σ) were calculated as described by Bevington (1969).

[b] Bone *et al.* (1989b).

[c] Bone *et al.* (1989a).

[d] Kettner *et al.* (1988).

[e] Dissociation constants were taken to be k_{off}/k_{on}.

[f] This inhibitor does not form a tetrahedral adduct with the enzyme (Bone *et al.*, 1989b).

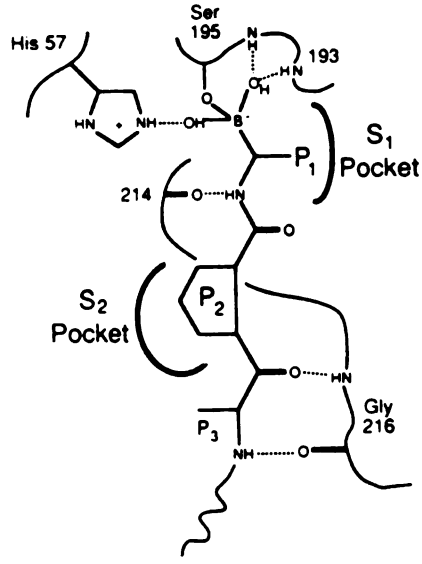


Figure 5.1 Schematic of the interactions between α-lytic protease and inhibitors. Six intermolecular hydrogen bonds stabilize the complex, as do electrostatic interactions between the negatively charged boronic acid and the positively charged histidine and hydrophobic interactions in the S_1 and S_2 specificity pockets. The primary specificity poscket of α-lytic protease (S_1) is lined by the side chain of residues Met 192, Met 213, and Val 217A.

In this work we seek to understand how these mutant proteases are able to expand to accommodate side chains such as Leu and Phe in relatively small specificity pockets and how side chains as diverse as Ala and Phe can fit well into the same specificity pocket with little loss in activity. To answer this question we have used x-ray crystallography to determine the high-resolution structures of 10 enzyme-inhibitor complexes between the mutant α-lytic proteases and a series of inhibitory peptide boronic acids. These inhibitors provide excellent mimics of the transition state for substrate hydrolysis (Bone *et al.*, 1987, Bone *et al.*, 1989b), and systematic variation of the boronic acid side chain allows structural analysis to parallel kinetic studies on enzyme specificity. When combined with previously-determined crystal structures, the resulting data suggest a model for how a series of substrates interact with wild type and mutant enzymes, how interactions differ between these enzymes, and how each enzyme changes in structure to accommodate each inhibitor.

Materials and Methods

Cloned wild type and mutant α-lytic proteases were purified from culture filtrates of *E. coli* as previously described (Silen *et. al.*, 1989; Hunkapiller *et al.*, 1973; Whitaker, 1970) and migrated as single bands when subjected to low pH native polyacrylamide gel electrophoresis (Hames & Rickwood, 1981). sAAP-Gly-*p*-Na, and sAAP-Norleu-*p*-Na were purchased from Bachem Inc. mAAP-boroAla, mAAP-boroVal, mAAP-boroNorleu, mAAP-boroLeu, mAAP-boroPhe were synthesized as previously described (Kettner *et al.*, 1988; Kettner & Shenvi, 1984; Bone *et al.*, 1989b).

Values of k_{cat} and K_M for these substrates were determined from double reciprocal plots of initial reaction velocity as a function of substrate concentration and errors estimated as described by Bevington (1969). Reactions, initiated by the addition of enzyme, contained substrate and 0.1 M Tris-Cl (pH=8.0) and were monitored continuously at 410 nm for the production of *p*-nitroaniline ($\Delta\varepsilon$ =8860 M^{-1} cm^{-1}; Hunkapiller *et al.*, 1976). The K_i values for all inhibitors, except those noted below, were determined by monitoring reaction velocity as a function of inhibitor concentration at substrate concentrations far below the K_M ([S]=K_M/20). Under these conditions the K_i value is the reciprocal slope of a plot of V_O/V_i as a function of inhibitor concentration, [I], where V_O is the initial reaction velocity in the absence of inhibitor and V_i is the initial reaction velocity in the presence of inhibitor at concentration [I].

K_i values for the binding of mAAP-boroVal, mAAP-boroNorleu, and mAAP-boroLeu by Met 192 -> Ala were calculated from the rate constants for association and dissociation of the complex, which were determined as described by Williams and Morrison (1979) for very slow tight-binding

inhibitors. These studies were done in the presence of 0.1 mg/ml polylysine (Ave. MWt. 300000, Sigma Chemical Co.) to stabilize the enzyme at very low concentrations ([E] <4 nM). Values of K_i determined from the ratio of k_{off} to k_{on}, assuming that enzyme and inhibitor combine slowly to form the final E-I complex, were checked by calculating K_i values from the extent to which preformed E-I complexes reverse to yield free enzyme. For the binding of mAAP-boroVal and mAAP-boroPhe (Bone *et al.*, 1989a; Bone unpublished) by Met 192 -> Ala these values matched k_{off}/k_{on}. However, for the binding of mAAP-boroNorleu and mAAP-boroLeu by Met 192 -> Ala there were discrepancies of factors of 10 and 3 in the apparent K_i values, suggesting that binding was tighter than the ratio of k_{off} to k_{on}. Such discrepancies may be indicative of an even slower isomerization step that follows slow association of the enzyme and inhibitor with apparent k_{off} values representing inhibitor dissociation from the relatively small portion of E-I complex present in unisomerized form. Therefore, the apparent K_i values for these two complexes should be considered upper estimates of the true K_i values.

Crystals of α-lytic protease and mutants were grown from 1.3 M lithium sulfate containing tris-sulfate (20 mM, pH 8.0) at ambient temperature (Bone *et al.*, 1989a; Bone *et al.*, 1987; Brayer *et al.*, 1979b). To prepare enzyme-inhibitor complexes, inhibitor solutions (0.1-0.5 M) in water were added in small aliquots (0.25-1.0 μl) to vapor diffusion droplets (5 μ l) containing 1-3 crystals of α-lytic protease. Generally, several aliquots of inhibitor solution were added to the same crystal drop, allowing 10-24 hours equilibration between additions, to achieve the final inhibitor concentration (Table 5-II). Protease crystals remained undamaged by this procedure.

Table 5-II: Crystallography of α-Lytic Protease-Inhibitor Complexes

STRUCTURE[a]	[I][b] (mM)	Res.[c] (Å)	R^d	% Rfls.[e] $I \geq 3\sigma_I$	RMS Deviations[f] Bond	Angle
Wild type + mAAP-boroLeu	50	2.20	0.129	92	0.017	0.053
MA192 + mAAP-boroAla	150	2.12	0.136	89	0.017	0.054
MA192 + mAAP-boroVal	40	2.25	0.127	87	0.016	0.053
MA192 + mAAP-boroNorleu	50	2.15	0.138	91	0.017	0.052
MA192 + mAAP-boroLeu	25	2.10	0.136	87	0.018	0.055
MA213 + mAAP-boroAla	75	2.13	0.134	93	0.020	0.045
MA213 + mAAP-boroNorleu	25	2.10	0.132	92	0.018	0.053
MA213 + mAAP-boroLeu	75	2.05	0.136	82	0.019	0.057
MA213 + mAAP-boroPhe	100	2.25	0.132	82	0.020	0.045

[a] The "boro" prefix indicates that inhibitors are the α-amino boronic acid analog of the corresponding amino acid.
[b] Final inhibitor concentration durring crystal soaking.
[c] Highest resolution of the data.
[d] Conventional crystallographic *R*-factor.
[e] % of the data collected with intensities greater than 3 σ_I; all of the data were used in structure determination and refinement.
[f] Root mean square deviations from ideal bond and angle distances.

Data from crystals of enzyme-inhibitor complexes were collected from single crystals using either Syntex P21 or Rigaku AFC5 automated diffractometers equipped with graphite monochrometers (Stroud et al., 1974; Wycoff et al., 1967). Obtaining high resolution data generally required that diffraction data from two or three crystals be merged. To date, all mutants and enzyme-inhibitor complexes have crystalized isomorphously in the space group $P3_221$, showing only small deviations from the unliganded wild type cell dimensions (less than 0.15%; $a=b=66.3$ Å $c=80.1$ Å $\alpha=\beta 90°$ $\gamma=120°$). The intensities of seven check reflections were monitored in order to correct for crystal decay, which was less than 25% over the course of data collection. Corrections were also made for absorption, the effect on intensities described by Lorentzian geometry, and polarization using standard methods and backgrounds according to Krieger et al. (1974). Data were collected by ω scan in shells of 2θ with the scan rate adjusted so that at least 81% of the reflections had intensities greater than $3\sigma_I$ over the entire data set (Table 5-II).

Initial difference Fourier maps were computed using structure factors and phases calculated from the refined coordinates of the complex between α-lytic protease and mAAP-boroVal (Bone *et al.*, 1989) from which the inhibitor had been removed. Maps were inspected and the inhibitors were placed using the interactive graphics package FRODO (Jones, 1982) after which the coordinates were refined by the stereochemically restrained least-squares algorithm of Hendrickson and Konnert (1981), modified for multiple occupancy refinement (Smith *et al.*, 1988), adapted for use on the FPS 264 array processor (Furey, 1984) and further modified by us. The final overall crystallographic *R* factors for the comparison of observed and calculated structure factors are listed in Table 5-II. Average α-lytic protease temperature

factors were 11 $Å^2$ while inhibitor temperature factors averaged 16.5 $Å^2$, increasing from the P_1 to the P_4 residue. Each structure also contained 1 sulfate and 160-180 molecules of solvent. Surface area calculations (Table 5-III) were done using the molecular surfacing program MS(Connolly, 1983a,b). The final refined coordinates for the structures reported in this work have been submitted to the Brookhaven Protein Data Bank.

Table 5-III: Structural Characterization of Wild-Type and Mutant α-Lytic Proteases

Inhibitor P_1 Residue[h]	Ave. H-Bond Length[a] (Å)	Buried Surface Area[b] Total (Å^2)	Buried Surface Area[b] Hydrophobic (Å^2)	Cavity Volume[c] (Å^3)	RMS Mainchain Deviation[d] (Å)	Torsional Shift From Free Enzyme $\Delta\Psi217 - \Delta\Phi217A$[e]
				Wild Type		
Ala[f]	2.79	500	300	58	0.121	-6.5
Val[f]	2.88	550	350	18	0.143	0.0
Ile[f]	2.92	540	340	3.8	0.165	7.2
NorLeu[f]	2.95	540	340	10	0.159	53
Leu	2.96	540	340	16	0.128	-1.0
				MA192		
Ala	2.86	490	300	69	0.123	-6.0
Val	2.89	530	340	29	0.130	-1.0
NorLeu	2.87	650	420	58	0.132	-5.3
Leu	2.92	550	350	36	0.136	3.7
Phe[g]	2.89	660	430	42	0.143	11.4
				MA213		
Ala	2.91	480	290	97	0.119	3.4
Val[g]	2.93	560	360	36	0.116	5.3
NorLeu	2.90	610	380	53	0.139	3.8
Leu	2.90	550	350	27	0.130	7.2
Phe	2.92	620	400	34	0.163	46

[a] Average of the 6 hydrogen bonds stabilizing the enzyme inhibitor complex: Ser 195 N-boro P_1 O_1; Gly 193 N-boro P_1 O_1; His 57 $N\varepsilon_2$-boro P_1 O_2; Ser 214 O-boro P_1 N; Gly 216 N-Ala P_3 O; Gly 216 O-Ala-P_3 N (Bone *et al.*, 1987).

[b] Buried Surface area calculated according to Connoly (1983a,b).

[c] Cavity volumes were calculated by subtracting the sum of the volumes of the sidechains projecting into the specificity pocket (P_1, 192, 213,217A) from the volume of the specificity pockets (Bone & Agard, 1991). The volume of a molecule of solvent was subtracted from the cavity volume for the MA213 complex with mAAp-boroAla.

[d] RMS deviation of all mainchain atoms (N, Cα, C, O) from the unliganded strucutres (Fujinaga *et al.*, 1985; Bone *et al.*, 1989a).

[e] Conformational change in the region of Val 217A expressed in degrees as the deviation in torsional angles ($\Psi217_L - \Psi217_U$) - ($\Phi217A_L - \Phi217A_U$) where L and U refer to the liganded and unliganded forms of the enzyme.

[f] Structures determined by Bone *et al.* (1989a).

[g] Structures determined by Bone *et al.* (1989b).

[h] Inhibitors are mAAP-boroX.

Results

The experiments undertaken in this report represent an effort to directly correlate structure with function. Towards this end, the structures of enzyme-inhibitor complexes have been determined that correspond to each enzyme substrate complex (except Gly and Met) during the transition state. Previously determined specificity profiles have been expanded to include two new substrates (with Gly and Norleu in P_1). The question of how well inhibitors mimic the transition state is addressed and the general features and energetics of inhibitor binding are examined followed by brief descriptions of the structures.

Kinetics. The specificity profiles of wild type and mutant α-lytic proteases (Bone *et al.*, 1989a) were expanded to include values of k_{cat}, K_m and k_{cat}/K_m for sAAP-Gly-*p*-Na and sAAP-Norleu-*p*-Na (Table 5-I). The effect of reducing the size of the P_1 side chain from Ala to Gly is a reduction in activity by a factor of 310 for Met 192 -> Ala, 64 for the wild type enzyme and 7.7 for Met 213 -> Ala. Met 192 -> Ala stabilizes the transition state of the substrate with Ala in the P_1 position by 3.4 Kcal/mole more than it does the substrate with a P_1 Gly, a value approaching the highest incremental binding energy observed for selection of a methyl group over a hydrogen (Fersht, 1985). The potent Ala/Gly selectivity of Met 192 -> Ala can not be simply a function of the composition or volume of the specificity pocket, else Met 213 -> Ala or the wild type enzyme would exhibit better Ala/Gly selectivity. Structural analysis of enzyme-inhibitor complexes is necessary in order to understand what structural properties give rise to such good discrimination between methyl and hydrogen substituents.

Because Norleu boronic acids have been used as analogs of Met in inhibitor studies, the activity of α-lytic proteases towards sAAP-Norleu-*p*-Na was examined. Though the volumes of Norleu, Met and Leu side chains are approximately the same (Zamyatnin, 1972), the activity of wild type α-lytic protease towards substrates with these side chains differ considerably (Table 5-I). For example, wild type α-lytic protease has 24 fold higher activity towards sAAP-Met-*p*-Na than towards the corresponding Norleu substrate. Because the activity of the enzyme towards sAAP-Norleu-*p*-Na fits the general downward trend in activity as a function of increasing P_1 side chain bulkiness (Table 5-I) and because the activity towards sAAP-Met-p-Na does not, it appears that sAAP-Met-*p*-Na is an anomalously good substrate for the enzyme. One possible explanation for these results is that C-S-C bond angles are easier to deform, allowing the Met side chain to avoid steric contacts that reduce activity towards the substrate with Norleu in the P_1 position. In contrast, both mutant enzymes prefer slightly the Norleu substrate over the Met substrate, with Met 192 -> Ala and sAAP-Norleu-*p*-Na having the highest activity of any α-lytic protease-substrate pair ($k_{cat}/K_M = 10^6$). The Met-NorLeu anomaly is abolished in the mutants because these substrate side chains are expected to bind in more extended conformations (see below) that do not require distortion of the C-S-C bond angle. A clear implication of these results is that caution must be observed in extending to Met substrates those conclusions drawn from the structures of enzyme-inhibitor complexes with Norleu boronic acids.

K_i values for inhibitors corresponding in structure to substrates are listed in Table 5-I. For a homologous series of structurally varying transition state analogs and substrates, a linear relationship is expected in plots of $\log[K_i]$ vs. $\log[k_{cat}/K_m]$ (Bartlett & Marlowe, 1983; Westerik & Wolfenden, 1972). For α-lytic protease such a linear relationship is observed between inhibition and

activity (Fig 5.2) upon variation of both the enzyme (mutants) and substrate (alternate P_1 amino acids). The correlation coefficient for the least squares analysis of 19 enzyme-inhibitor and enzyme-substrate pairs, which include values for a Val 217A->Ala mutant (Bone *et al.*, 1991a), is 0.87 with a slope of -0.70. If the two most deviant data points, corresponding to interaction of the wild type enzyme and Met 192 -> Ala with substrate/inhibitor pairs with a Val side chain in the P_1 position (Fig 5.2, arrows), are omitted the correlation coefficient increases to 0.92 with a slope of -0.69. The slope of -0.7 indicates that enzyme-substrate interactions are more sensitive to changes in structure than corresponding enzyme-inhibitor interactions. Correlations with similar slopes have been observed for elastase (peptide aldehydes; Thompson, 1973) and chymotrypsin (trifluoroketones; Brady & Abeles, 1990). Transition state analog inhibitors of serine proteases appear to be very good but not exact analogs of the transient, high energy transition state for substrate hydrolysis. An alternate possibility is that for some enzyme-substrate pairs the rate determining step for the reaction may have changed. Inhibitors might not be sensitive to subtle differences in the way in which such substrates might interact with the enzyme.

General Aspects of Inhibitor Binding. Difference electron density maps clearly showed that all inhibitory boronic acids were bound by the enzyme at a single site with high occupancy. Maps calculated with Fourier coefficients ($2|F_O|-|F_C|$) showed continuous electron density stretching between the active site Ser (195 $O\gamma$) and the boron of the inhibitors, indicating that in each complex a covalent adduct had formed with the active site serine. This was corroborated by the bond lengths found on refinement. On the basis of the shape of the electron density about the boron and bond angles after refinement, the geometry of the adducts were slightly distorted from

tetrahedral[4] (Figs 5.3-5.5). No electron density is observed after the P_4 Ala residue for any of the inhibitors, indicating that the methoxysuccinyl group is disordered.

In qualitative terms, the intermolecular interactions that stabilize the E-I complex are essentially identical to those that stabilize the complex formed between α-lytic protease and bAP-boroVal (Fig 5.1; described in detail in Bone *et al.*, 1987). The inhibitors make the same 6 hydrogen bonds with the enzyme (2 in the oxyanion binding pocket with the amide groups of Ser 195 and Gly 193, 1 with His-57 and 3 with main chain elements of residues 214-216) and have nearly the same interatomic distances between hydrogen bonding atoms (Figs 5.3-5.5). These inhibitors also make van der Waals contacts with the side chains of Met-192, Met-213 and Val-217A, and the main chains of residues 214-216 and 192-193 in the primary specificity pocket and with the side chains of Tyr-171, Phe-94, and His-57 in the P_2 binding site. Conformational changes outside of the specificity pocket, most notably in the S_2 binding site where residues 169-171 adjust away from the enzyme to accommodate the inhibitor, are very similar to those that have been characterized previously (Bone *et al.*, 1987; Bone *et al.*, 1989b). The combination of hydrogen bonds, electrostatic interactions between the positively charged histidine and the negatively charged boron (Bachovchin *et al.*, 1988), and the removal of hydrophobic surfaces in the P_1 and P_2 specificity sites stabilize these complexes and presumably also the transition state for substrate hydrolysis.

Where this series of structures differ is in the details of the interactions of the P_1 side chain with the specificity pocket and how these interactions affect the energetics of enzyme-inhibitor complex formation. In addition to qualitative descriptions of how the mutations have affected interactions with inhibitors (Figs 5.3-5.5), it is possible to analyze the structures in more

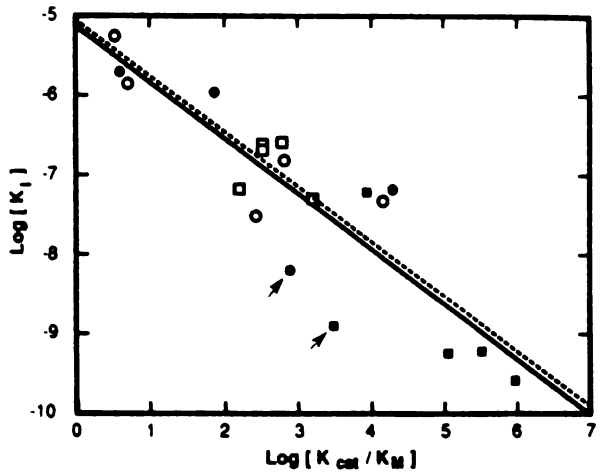


Fig. 5.2 Plot log k_i vs. log (k_{cat}/K_m) for wild-type and mutant α-lytic proteases: Closed circles, wild-type (Bone et al., 1989a); open circles, VA217A (Bone et al., ms. in preparation); closed squares, MA192 (Bone et al., 1989a); open squares, MA213 (Bone et al., 1989a). Substrates/inhibitors used were sAAP-X-p-Na/mAAP-boroX, where X is Ala, Val, Ile (wild-type only), Norleu, Leu, or Phe (except wild-type) for each enzyme. The slope of the line is 0.7 with a correlation coefficient of 0.87. Without the arrowed data points, corresponding to substrates and inhibitors with Val in the P_1 position, the slope is unchanged and the correlation coefficient rises to 0.92.

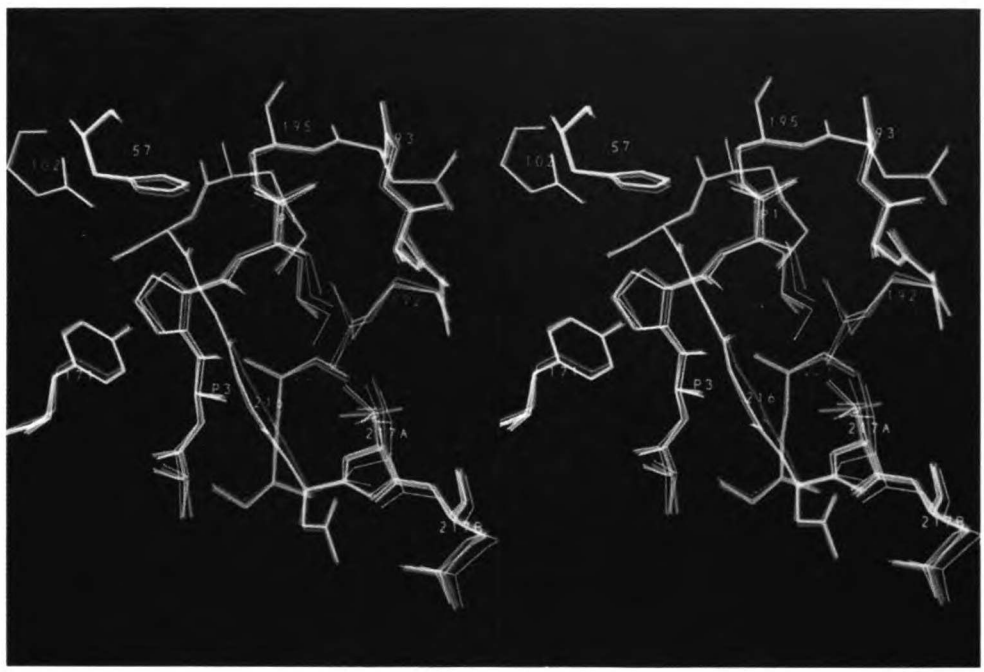


Fig. 5.3 Stereo drawing of the superimposed structures of the specificity posket regions of wild-type α-lytic protease complexes with each of five boronic acid inhibitors: purple, uncomplexed; orange, mAAP-boroAla; green, mAAP-boroVal; red, mAAP-boroIle; yellow, mAAP-boroNorleu; and blue, mAAP-boroLeu. The enzyme portion of each complexed structure was superimposed on the uncomplexed enzyme so as to minimize the rms deviation between α-carbon positions. Alternate P_1 side chains are accommodated via adjustments in the region of Val 217A and with alternate conformations of Met 192.

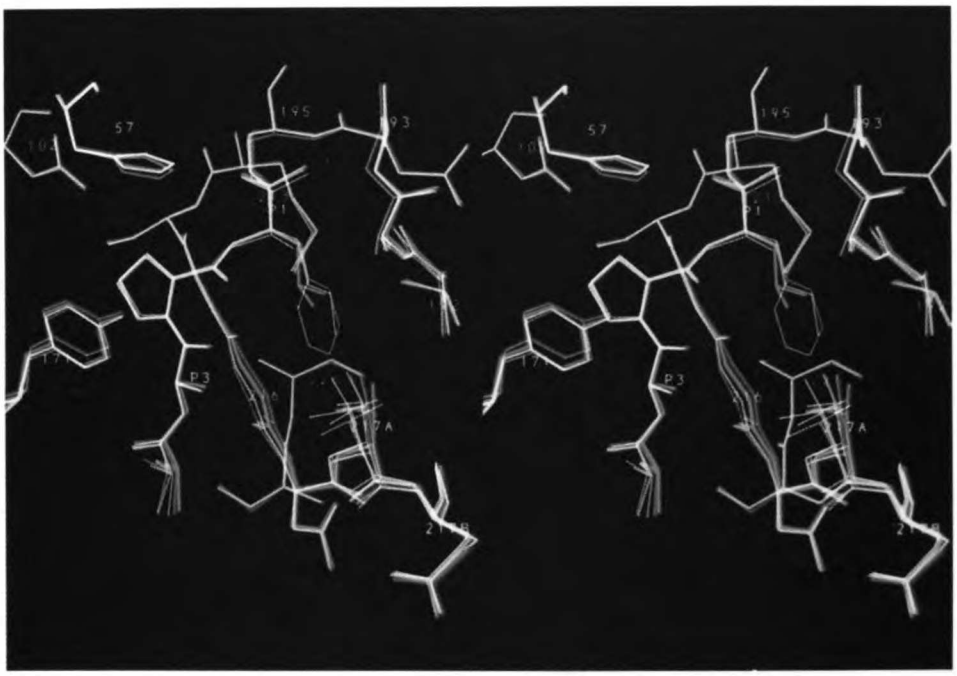


Fig. 5.4 Stereo drawing of the superimposed structures of the specificity pocket regions of MA192 complexes with each of five boronic acid inhibitors: purple, uncomplexed; orange, mAAP-boroAla; green, mAAP-boroVal; yellow, mAAP-boroNorleu; blue, mAAP-boroLeu; and red, mAAP-boroPhe. The enzyme portion of each complexed structure was superimposed on the uncomplexed enzyme so as to minimize the rms deviation between α-carbon positions. Alternate P_1 side chains are accommodated via adjustments in the region of Val 217A and Gly 216.

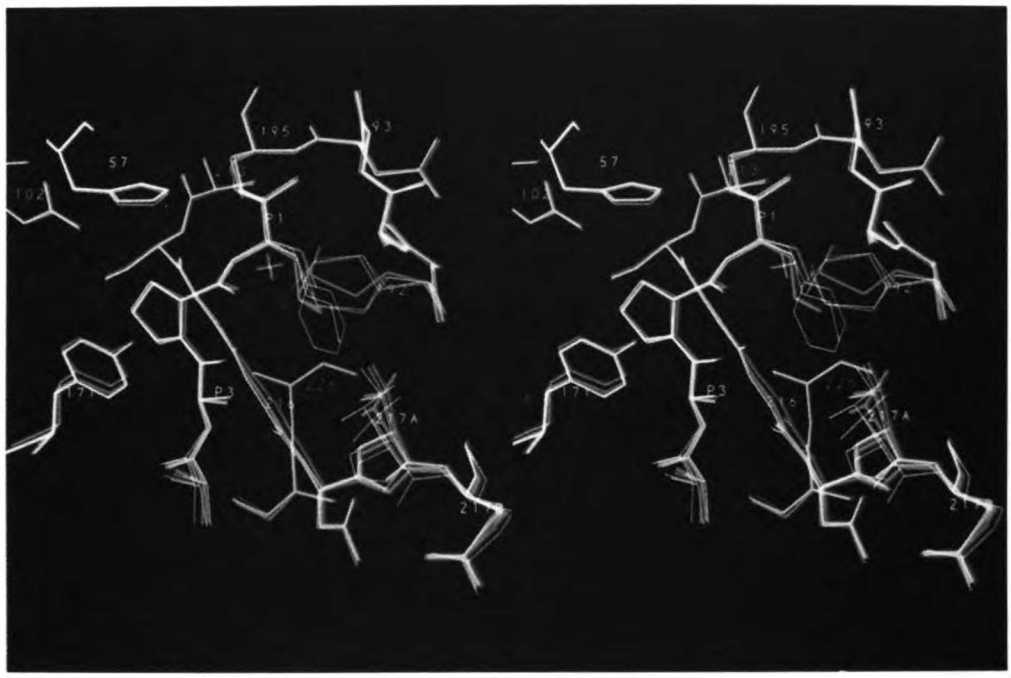


Fig. 5.5 Stereo drawing of the superimposed structures of the specificity pocket regions of MA213 complexes with each of five boronic acid inhibitors: purple, uncomplexed; orange, mAAP-boroAla; green, mAAP-boroVal; yellow, mAAP-boroNorleu; blue, mAAP-boroLeu; and red, mAAP-boroPhe. The enzyme portion of each complexed structure was superimposed on the uncomplexed enzyme so as to minimize the rms deviation between α-carbon positions. Alternate P_1 side chains are accommodated via adjustments in the region of Val 217A and with alternate conformations of Met 192.

quantitative terms. For each high-resolution structure we have calculated (Tables 5-III and 5-IV) the intermolecular hydrogen bond lengths (and the average), the total hydrophobic surface area buried upon complex formation (Connolly, 1983a,b), the sizes of the void volumes created in the specificity pockets upon inhibitor binding (cavity volume, see Table 5-III; Connolly, 1985; Bone & Agard, 1991) and the magnitude of the main chain conformational adjustment in the region of Val 217A (measured by the sum of the changes in torsional angles Ψ_{217} and Φ_{217A} relative to the unliganded enzymes; Bone *et al.*, 1989b; Bone & Agard, 1991).

Structural Parameters and Energetics. The structural parameters reported in Table 5-III provide a more detailed understanding of how enzyme-inhibitor interactions have changed as a result of mutation by quantifying the energetic contributions of the changes in interactions. Changes in parameters such as average hydrogen bond length have predictable energetic effects which indicate whether observed changes in structure enhance or degrade the stability of enzyme inhibitor complexes and by what quantity. Hydrogen bonds can be considered primarily electrostatic interactions, the energy of which will fall off as the bonds become stretched or bent (Schulz & Schirmer, 1979). If hydrogen bond is stretched by 0.3 Å or bent by 20° the energy of the interaction will decrease by approximately 10-15 % (Pimentel & McClellan, 1960; Mitchel & Price, 1990; Reed *et al.*, 1986; Cybulski & Scheiner, 1989). For a hydrogen bond between neutral partners the intrinsic energy of the interaction is 6 Kcal/mole (Weiner et al., 1984) and a reduction of 10% would decrease inhibitor affinity by 0.6 Kcal/mole. However, uncertainty in atomic positions, and therefore hydrogen bond lengths, due to experimental error and the resolution limit of these studies is estimated to be 0.12 Å (Bone &

Table 5-IV:Intermolecular Hydrogen Bond Distances in α-Lytic Protease-Inhibitor Complexes[a]

HYDROGEN BOND	ENZYME: WILD TYPE — Inhibitor Side Chain[a]					ENZYME: MA192 — Inhibitor Side Chain[a]					ENZYME: MA213 — Inhibitor Side Chain[a]				
	Ala[b]	Val[b]	Ile[b]	Norleu[b]	Leu	Ala	Val	Norleu	Leu	Phe[c]	Ala	Val[c]	Norleu	Leu	Phe
	(Å)					(Å)					(Å)				
Ser-195 N–P_1 O_1	2.97	2.97	3.03	3.14	3.20	3.00	3.00	3.03	3.02	2.92	3.01	3.04	3.03	3.11	3.06
Gly-193 N–P_1 O_1	2.45	2.59	2.55	2.54	2.71	2.46	2.63	2.59	2.51	2.60	2.46	2.44	2.49	2.41	2.48
His-57 $N_{\epsilon 2}$–P_1 O_2	2.65	2.62	2.76	2.65	2.47	2.78	2.64	2.64	2.77	2.74	2.80	2.76	2.71	2.78	2.62
Ser-214 O–P_1 N	2.93	3.11	2.99	3.20	3.21	2.98	3.20	3.05	3.23	3.09	3.19	3.20	3.14	3.14	3.11
Gly-216 N–Ala-P_3 O	2.90	2.93	3.08	3.12	3.11	2.96	2.95	3.01	3.04	3.07	3.09	3.06	2.97	3.06	3.14
Gly-216 O–Ala-P_3 N	2.87	3.03	3.13	3.06	3.06	2.96	2.95	2.92	2.92	2.90	2.93	3.11	3.02	2.92	3.09

[a] P_1 Side chain of mAAp-boroX, where the "boro" prefix indicates that the inhibitor is the α-amino boronic acid analog of the corresponding amino acid.

[b] Bone *et al.* (1989b).

[c] Bone *et al.* (1989a).

Agard 1991). In support of this value, difference Fouriers calculated using observed structure factors from different enzyme-inhibitor complexes and calculated phases clearly indicate positional adjustments of 0.2 Å. We consider changes in hydrogen bond lengths of 0.2 Å to be significant and well defined by our high resolution and highly refined data.

Based on free energies of transfer, removing hydrophobic surfaces from solvent should contribute 25 cal/$Å^2$ to the free energy of inhibitor binding (Richards *et al.*, 1977); removing an extra 55 $Å^2$ of hydrophobic surface from solvent should improve catalysis or binding 10 fold. The creation of void volumes upon complex formation is an unfavorable consequence of binding which will reduce binding by approximately 60 cal/$Å^3$ (Rashin *et al.*, 1986). The amount of energy required to create a water sized cavity, 11 $Å^3$, is estimated to be 0.66 Kcal/mole and should decrease activity by a factor of nearly 3. Conformational changes resulting from inhibitor binding require energy and will detract from inhibitor affinity, although the amount of energy is uncertain.

Native complex. In a previous study we examined how peptide boronic acids with Ala, Val, Ile and Norleu side chains in the P_1 position were accommodated by the wild type enzyme and found that main chain flexibility in the region of Val 217A, involving rotations of Ψ_{217} and Φ_{217A}, was the primary conformational adjustment (Bone *et al.*, 1989b). The magnitude of the conformational changes increase and intermolecular hydrogen bonds stretch as the size of the side chain projecting into the specificity pocket increases (Bone *et al.*, 1989b). In this work we have extended those studies by determining the structure of the complex between α-lytic protease and mAAP-boroLeu. Rather than an adjustment in the main chain conformation of

residues 217-217A, the side chain Val 217A rotates by 120° about the Cα-Cβ bond to remove the Val methyl group from the pocket and allow the P_1 Leu side chain into the specificity pocket (Fig 5.3). As a consequence of the poor complementarity between the Leu side chain and the specificity pocket, Met 192 is pushed deeper into the pocket and C_ε occupies a position occupied by S_δ in the other complexes (Fig 5.3). The side chain adjustment leaves Cε and Cβ in an unfavorable eclipsed conformation (χ_3 = 15°) which must detract significantly from the stability of the complex; the energetic cost of maintaining a fully eclipsed Met side chain conformation would be 3-5 Kcal/mole (Jainin *et al.*, 1978). Adjustment in the position of Met 192 forces the residues surrounding the Met, such as Ser 226, towards the body of the protein. Positioning of the inhibitor in the active site is also suboptimal as a result of the poor complementarity. The C_α and C_β of the boronic acid residue are forced away from the more optimal positions that they occupy in complexes having Ala or Val in the P_1 position (Fig 5.3). In addition, the hydrogen bonds between the enzyme and inhibitor become stretched[5,6] to nearly the same extent as in the complex with mAAP-boroNorleu (Tables 5-III and 5-IV). Despite these adjustments, no more hydrophobic surface is removed from solvent than in the other complexes (Table 5-III).

Activity towards the substrate with Leu in the P_1 position is reduced greatly because hydrogen bonds are not optimal, the Leu side chain is not removed completely from solvent and because transition state stabilization energy must be diverted to drive the Val 217A and Met 192 conformational changes. Because three fundamentally different conformational adjustments (the main chain shift in the region of Val 217A, the rotation of the Val 217A side chain and the rotation of the side chain of Met 192) are used separately or in combination to accommodate inhibitor side chains in the specificity pocket,

the energetic cost of these changes must be on roughly the same scale. If this were not the case, only the lowest energy conformational change would be observed to accommodate substrates. Substrates and inhibitors with large side chains in the P_1 position are selected against by the wild type enzyme because intermolecular hydrogen bonds are stretched, the conformation of the enzyme must be changed, and there is no compensating gain in hydrophobic interactions. Partially compensating these unfavorable results of binding is the reduction in the size of the cavity created when larger P_1 side chains interact with the specificity pocket.

Accommodation of side chains by Met 192 -> Ala. Met 192 -> Ala displays broad specificity, yet maintains high activity (Bone *et al.*, 1989a). Inhibitors with increasingly large P_1 side chains are accommodated in the specificity pocket of Met 192 -> Ala in the same manner as they are in the wild type specificity pocket. That is, the main chain of Val 217A shifts away from the center of the pocket and/or the side chain of Val 217A rotates by 120° to remove the Val methyl group from the pocket and to allow large substrate side chains access (Fig 5.4). The position of Gly 216 also adjusts in response to the nature of the P_1 side chain, although it does not in either wild type or Met 213 -> Ala complexes with inhibitors (Figs 5.3-5.5). From this observation, it appears that one role of Met 192 is to restrict the position of Gly 216 and that its absence in Met 192 -> Ala is partially responsible for the ability of the specificity pocket to shrink.

In the complex of Met 192 -> Ala with mAAP-boroAla, C_β of the P_1 Ala moves further into the specificity pocket than in other complexes, the sides of the pocket (residues 192-193, 215-216) which have contracted in the free mutant enzyme (Bone *et al.*, 1989a) remain contracted and Val 217A moves

further into the pocket. The net result is minimization of the size of the remaining cavity, reduction of the cross section of the cavity to less than the diameter of a water molecule and exclusion of solvent. There is remarkably little increase in the cavity volume in the complex of Met 192 -> Ala with mAAP-boroAla compared to the wild type complex with the same inhibitor (Table 5-III). Although a Met side chain has been removed from the pocket, with a volume of 80 $Å^3$, the increase in cavity size is only 11 $Å^3$ providing an indication of the extent to which the specificity pocket has shrunk. This shrinkage appears to be the main reason for the maintenance of high activity towards substrates with small P_1 side chains.

As the size of the P_1 side chain becomes larger, the pocket expands in volume (Bone & Agard, 1991) through conformational adjustments in the region of Val 217A and Gly 216 (Table 5-III). Smaller structural distortions are required to accommodate large side chains in the specificity pocket in Met 192 -> Ala-inhibitor complexes than in wild type complexes (Table 5-III). Also in contrast with the wild type enzyme, there is no tendency for intermolecular hydrogen bonds to become stretched as the P_1 side chain becomes larger and the amount of hydrophobic surface area removed from solvent increases dramatically (Tables 5-III and 5-IV). In complexes with Met 192 -> Ala, boronic acids with Norleu and Phe in the P_1 position bury 70 and 80 $Å^2$ more hydrophobic surface than the wild type complex with mAAP-boroVal and 120 and 130 $Å^2$ more than the wild type complex with mAAP-boroAla. This improvement in hydrophobic interactions could enhance binding or catalysis up to 200 fold. Improvements in activity and affinity on this scale are observed for the interaction of the mutant with either inhibitors or substrates (Table 5-I). For complexes of Met 192 -> Ala with inhibitors that have large P_1 side chains, the cavity volume decreases, but never to the extent of the wild

type enzyme. Because these alterations in cavity volume are quite small for Met 192 -> Ala, they are unlikely to significantly influence the specificity profile. Thus, broad specificity results from improved hydrogen bonds, increased hydrophobic interactions and attenuated conformational adjustments, all of which lead to high activity on substrates with large P_1 side chains. The ability of the specificity pocket to shrink provides for high activity on substrates with small P_1 side chains.

On the basis of the structure of the complex between mAAP-boroNorleu and Met 192 -> Ala it appears that the energetics of the conformational adjustments in the region of Val 217A are altered as compared with the wild type enzyme. In this structure, Val 217A occupies two conformations: in one conformation the Val 217A side chain is rotated into the specificity pocket and in the other the side chain is rotated out of the pocket (Fig 5.4). The two conformations, each at about 50% occupancy, are clearly indicated by patterns of positive and negative electron density in initial difference maps and in difference maps calculated after refinement with one or the other of the conformations at full occupancy. The conformation of the main chain (Φ_{217} = 76, Ψ_{217A} = -64) is almost exactly the same as in the wild type complex with the Ala boronic acid (Φ_{217} = 74, Ψ_{217A} = -66). The Val side chain rotation is not observed in wild type complexes with borolle and boroNorleu inhibitors which require a large and costly main chain shift in order to be accommodated (Bone *et al.*, 1989b), but is observed here where the main chain conformation seems optimal. From these observations, it appears that the energy required to rotate Val 217A out of the specificity pocket is less in the mutant that in the wild type enzyme. This may not be surprising since elimination of Met 192 from the specificity pocket by mutation eliminates hydrophobic interactions between

Val 217A and Met 192 that presumably participate in the stabilization of the native conformation.

Accommodation of side chains by Met 213 -> Ala. Met 213 -> Ala exhibits almost no selectivity for the P_1 side chain of substrates (Bone *et al.*, 1989a) or inhibitors and the activity of the enzyme has been reduced by approximately an order of magnitude relative to the wild type enzyme. As the P_1 side chain increases in size, the same combination of main chain flexibility in the region of Val 217A and rotation of the Val 217A side chain out of the specificity pocket is responsible for accommodating the P_1 side chains of inhibitors (Fig 5.5). However, in contrast with Met 192 -> Ala, Met 213 -> Ala also uses 3 different Met 192 conformations to adjust the shape of the specificity pocket (Fig 5.5) to complement the shape of the P_1 inhibitor side chain. Despite this seeming ability to tailor the shape of the specificity pocket, the Met 213 -> Ala specificity pocket still shows poor complementarity to substrate side chains as evidenced by the tendency of the boronic acid $C\beta$ to be pushed away from the center of the specificity pocket as the P_1 side chain increases in size. For both Met 213 -> Ala and the wild type enzyme this tendency appears to be an indicator of poor P_1 side chain complementarity (Figs 5.3-5.5). The specificity pocket of Met 213 -> Ala appears to be shallow rather than deep and although it can tolerate large side chains, it cannot accommodate them as well as Met 192 -> Ala.

Solvent is sequestered in the active site in both the free mutant (Bone *et al.*, 1989a) and in the complex with the Ala boronic acid (Fig 5.5). Only a hydrogen bond with the ring nitrogen of Trp 147 stabilizes the molecule of solvent in the enzyme-inhibitor complex. This is the only α-lytic protease complex examined to date in which a molecule of solvent is observed to be

buried in the specificity pocket on inhibitor binding. Evidently the presence of Met 192 prevents the type of shrinkage and solvent exclusion that occurs in the Met 192 -> Ala complex with mAAP-boroAla from also occurring in Met 213 -> Ala complexes. Because the specificity pocket is unable to shrink, the void volume created in the complex of Met 213 -> Ala with mAAP-boroAla is 40 $Å^3$ greater than in the corresponding wild type complex. The increased cavity volume in this complex destabilizes the enzyme-inhibitor complex by an estimated 2.3 Kcal/mole (Rashin, Iofin & Honig, 1986) and is sufficient to explain the reduced activity and affinity of the enzyme towards Ala substrates and inhibitors.

There is no tendency for hydrogen bonds to become stretched as size of the P_1 side chain of inhibitors increases, although the average hydrogen bond distances appear to be slightly longer than either wild type or Met 192 -> Ala intermolecular hydrogen bond lengths (Table 5-III and 5-IV). Hydrophobic interactions improve as the boronic acid side chain increases in size, but not to the extent observed in Met 192 -> Ala complexes. When Norleu and Phe side chains interact with the Met 213 -> Ala specificity pocket, 40 and 30 $Å^2$ less hydrophobic surface is buried, equivalent to approximately 1 Kcal/mole less binding energy, than when inhibitors with these side chains interact with Met 192 -> Ala. This must partially account for the decreased activity of Met 213 -> Ala as compared with Met 192 -> Ala towards substrates with large P_1 side chains. Another factor limiting the activity and affinity of Met 213 -> Ala for substrates and inhibitors with large P_1 side chains is that side chain accommodation requires conformational changes nearly as large as in the wild type-mAAP-boroNorleu complex (Table 5-III; Fig 5.5). In addition, it must also take energy to alter the conformation of Met 192 in these complexes from its conformation in the free mutant enzyme.

Discussion

Quality of Structural Models. Qualitatively, the enzyme-inhibitor complexes studied in this work appear to be excellent analogs of high energy reaction intermediates (Bone *et al.*, 1987). However, the question of how closely these complexes resemble the actual high energy, anionic, tetrahedral intermediates that are formed during substrate hydrolysis should be addressed. The correlation between log [K_i] and log [k_{cat}/K_M], based on data for 19 enzyme-substrate pairs varying over 5 orders of magnitude in activity, is quite good (Fig 5.2; correlation coefficient = 0.87) and improves significantly (correlation coefficient = 0.92) if two outlying points (wild type and Met 192 -> Ala enzymes paired with Val in the P_1 position) are omitted . These two deviant points, the scatter of the observations about the line and the slope of the plot (-0.7) suggest that some aspects of the transition state are not mimicked properly by the inhibitors.

The breakdown in transition state analogy is a likely consequence of subtle differences between the charge distributions of the bound inhibitors (negative charge on boron) and the true high energy intermediates (negative charge on oxygen) which are reinforced by the steric consequences of accommodating branched side chains in the specificity pocket (collision with Met 213; Bone *et al.*, 1989b). Consequently, hydrogen bond lengths and electrostatic interactions are expected to be slightly different in substrate complexes during the transition state than in boronic acid complexes. Differences in interatomic distances of only 0.1-0.3 Å would translate into changes in interaction energy of 0.5-2.5 Kcal/mole (Cybluski & Scheiner, 1989; Mitchel & Price, 1990; Reed *et al.*, 1986; Schulz & Schirmer, 1979; Weiner *et al.*, 1984) and would be sufficient to account for the deviation of even the most

scattered data from the correlation between transition state stabilization and inhibitor affinity. Therefore, the structures of substrate complexes during the transition state are not expected to differ significantly from the structures of boronic acid complexes in surface area buried, the extent of conformational changes or in the sizes of void volumes created upon inhibitor binding. In addition, relative changes in hydrogen bond lengths in boronic acid complexes should be representative of changes that occur in the structures of the corresponding transition states for substrate hydrolysis. With these limitations in mind, conclusions regarding the specificity profiles for inhibitor binding by these three enzymes should also apply to the specificity profiles for substrate hydrolysis.

Structural Basis for Specificity. Comparison of the structures of the enzyme-inhibitor complexes studied in this work provide a qualitative picture of how each enzyme interacts with substrates (Figs 5.3-5.5). More detailed analysis of the energetic components of inhibitor binding (Tables 5-III and 5-IV) has lead to an increased understanding of the structural and energetic basis for the substrate and inhibitor specificity profiles for these three enzymes. However, it is important to note that no single factor correlates well with inhibitor affinity or substrate activity. Nor does specificity correlate with properties of the substrates as has been observed for some other proteases (Dorovska *et al.*, 1972; Estell *et al.*, 1986).

The wild type specificity profile (Bone *et al.*, 1989a), for either inhibitors or substrates, appears to be driven for the most part by steric exclusion and limited flexibility. The unfavorable consequences of allowing large side chains into the specificity pocket (stretched hydrogen bonds and

induction of large conformational changes, Table 5-III), appear to be the primary determinants of the specificity profile.

Mutation of Met 192 to Ala creates an enzyme with an extraordinarily broad specificity profile because the contribution of unfavorable interactions to binding have been reduced. Hydrogen bonds do not become stretched as the size of the P_1 side chain increases and the conformational changes required to accommodate larger side chains have been greatly attenuated (Tables 5-III and 5-IV). The energy required to drive rotation of the side chain of Val 217A out of the specificity pocket has been reduced by mutation. In addition, hydrophobic interactions improve significantly as the size of the P_1 side chain increases, enhancing inhibitor affinity and substrate activity. Activity towards substrates with small side chains is retained because of the ability of the specificity pocket to shrink in the absence of Met 192; thereby reducing the size of the residual cavity in the enzyme-substrate complexes sufficiently to exclude solvent.

The Met 213 -> Ala mutation leads to a flexible enzyme that, like Met 192->Ala, can accommodate large variations in P1 side chain size, but has lost activity as a result of mutation (Bone *et al.*, 1989a). There appears to be no dominant factor responsible for the reduced activity; rather a combination of factors are involved. The creation of large void volumes reduces activity towards P_1 Ala and Val substrates while poor complementarity, less than optimal hydrophobic interactions and costly conformational changes reduce activity towards substrates with large side chains. Thus, it seems that Met 213 plays a significant role in either directly or indirectly effecting the precise positioning of substrates or inhibitors for optimal interactions with the enzyme.

While the trend in enzyme-inhibitor affinity appears to correlate with trends in average hydrogen bond length and in the magnitude of side chain accommodating conformational changes for the wild type enzyme, it does not for either Met 192 -> Ala or Met 213 -> Ala. Similarly, buried hydrophobic surface area may be an indicator of relative affinity for Met 192 -> Ala but is not for either of the other enzymes. For Met 213 -> Ala, inhibitor affinity does not change dramatically as the P_1 side chain is varied, but large variations are observed in hydrophobic interactions, cavity volumes and conformational changes. The specificity profiles of these enzymes cannot be understood in terms of only one factor such as the number of hydrogen bonds or the hydrophobicity of the P_1 side chain of the substrate. Rather, each specificity profile results from the summation of contributions from factors promoting complex formation (hydrogen bonding, hydrophobic interactions, electrostatic interactions) and factors discouraging complex formation (conformational changes, steric clashes leading to hydrogen bond distortion, creation of void volumes).

Discrimination of CH_3- and H- by Met 192 -> Ala. The specificity profiles of each enzyme have been expanded to include values for a substrate with Gly in the P_1 position. Of the 3 enzymes, Met 192 -> Ala is most effective at selecting for Ala in preference to Gly at the P_1 position of substrates. Elimination of the methyl group destabilizes the transition state by 3.4 Kcal/mole for amide hydrolysis catalyzed by Met 192 -> Ala (Table 5-I) and approaches the highest discrimination between methyl and hydrogen substituents observed (Fersht *et al.*, 1980). The specificity pocket of Met 192 -> Ala shrinks considerably in the absence of ligands (Bone *et al.*, 1989a) and when accommodating an Ala side chain in the P_1 position (Fig 4). It seems

unlikely that it would be able to collapse further. The ability of the enzyme to select against a Gly substrate is therefore a consequence of both the rigidity of the specificity pocket, at least in the sense of not having the capacity to shrink any further, and its oversized nature relative to the preferred substrate side chain (Ala). In the Met 213 -> Ala active site the size of the specificity pocket does not shrink, which leads to the sequestering of solvent when interacting with the Ala substrate. This results in a dramatic decrease in the Ala-Gly selectivity ($\Delta\Delta G$ = 1.2 Kcal/mole), primarily through a reduction in activity towards the Ala substrate. Selectivity of the wild type enzyme is intermediate between Met 192 -> Ala and Met 213 -> Ala ($\Delta\Delta G$ = 2.5 Kcal/mole), perhaps because the specificity pocket of the wild type enzyme has some ability to shrink and reduce size of the cavity when interacting with the P_1 Gly substrate. These results suggest that optimum selectivity for a large nonpolar substrate substituent over a smaller one occurs when the specificity pocket is somewhat larger than the substituent that it selects for and cannot adjust to accommodate the smaller substituent.

Broad Specificity and Conserved Conformational Changes. Although mutation of either Met 192 or Met 213 to Ala results in an enzyme that can efficiently hydrolyze substrates having Met just prior to the scissile bond, both mutant proteases exhibit extraordinarily broad specificity. Broad specificity in these two mutants appears to be a primary consequence of conformational flexibility in the region of Val 217A. It is the ability of the main chain and side chain torsional angles in this and adjacent residues to adjust that allows the specificity pocket to expand dramatically in volume and accommodate side chains larger than Met or Norleu. In addition, adjustments in these residues participate in the shrinkage of the pocket that allows Met 192 -> Ala to retain complementarity with substrates having smaller side chains.

The changes in conformation responsible for the accommodation of substrates by the wild type specificity pocket are conserved in the mutant enzymes, although the energy required for these conformational transitions may differ from the wild type enzyme. That each enzyme uses basically the same repertoire of responses as the primary means of accommodating substrate side chains in the specificity pocket suggests that the behaviors of mutant enzymes should be predictable if the wild type enzyme is sufficiently well characterized in structural terms. Furthermore, the presence of this flexibility in the wild type enzyme must be a factor promoting broadened specificity. If the enzyme were less flexible in the region of Val 217A it should exhibit higher selectivity for Ala in the P_1 position.

Conformational flexibility in the region of Val 217A is very likely derived from the secondary structure of an eight residue Ω loop (Leszczynski & Rose, 1986) beginning with Val 217A. This Ω loop appears to be able to readily absorb conformational adjustments that occur in the immediately preceding residues. Remaining unclear is how much energy each of these conformational adjustments requires and which intramolecular interactions determine the energy cost. The changes in main chain torsional angles in the region of Val217A do not represent transitions from inherently low energy Φ and ψ angular values to high energy ones (Brant *et al.*, 1967), so the energetics of these changes must be the result of inter-residue interactions.

From the work presented here, it is now clear that the patterns of substrate specificity observed for α-lytic protease result not just from the direct interactions of side chains of Met 192, Met 213, and Val 217A with the P1 side chain of the substrate. Conformational changes are central participants in substrate selection and the energy requirements of these changes must be due in large part to interactions with surrounding residues. Therefore, the

observed specificity profiles can be seen to arise indirectly from interactions that occur between peripheral regions of the protein and the active site region. These interactions dictate the degree of flexibility available to the active site residues by setting the energetic price for changing the conformation of the enzyme. One of the remaining challenges is to understand the energetic and structural origins of the interactions which act to modulate active-site flexibility.

Acknowledgements

We acknowledge Dr. Robert M. Stroud for allowing us the use of his diffractometer, Jim Mace for providing assistance with the kinetic analysis. Funding was provided by the Howard Hughes Medical Institute and through an NSF Presidential Young Investigator grant.

References for Chapter Five: Structure-Function

Bachovchin, W. W., Wong, W. Y. L., Farr-Jones, S., Shenvi, A. B. and Kettner, C. A. 1988. Nitrogen-15 NMR spectroscopy of the catalytic-triad histidine of a serine protease in peptide boronic acid inhibitor complexes. *Biochemistry* 27: 7689-7697.

Bartlett, P. A., & Marlowe, C. K. 1983. Phosphonamidates as transition-state analogue inhibitors of themolysin. *Biochemistry* 22: 4618-4624.

Bauer, C.A., Brayer, G. D., Sielecki, A.R., & James, M. N. G. 1981. Active site of α-lytic protease: enzyme substrate interactions. *Eur. J. Biochem.* 120: 289-294.

Bevington, P. R. 1969. *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, pp 92-118.

Bone, R., Silen, J. L., & Agard, D. A. 1989a. Structural plasticity broadens specificity of an engineered protease. *Nature* 339:191-195.

Bone, R., Frank, D., Kettner, C. A., & Agard, D. A. 1989b. Structural analysis of specificity: α-lytic protease complexes with analogues of reaction intermediates. *Biochemistry* 28: 7600-7609.

Bone, R. & Agard, D. A. 1991. In preparation.

Bone, R., Madhani, H. and Agard, D.A. 1991a. In preparation.

Bone, R., Kettner, C. A. & Agard, D. A. 1991b. In preparation.

Bone, R., Sampson, N. S., Bartlett, P. A. & Agard, D. A. 1991c. *Biochemistry* 30: 2263-2272.

Bone, R., Shenvi, A. B., Kettner, C. A., & Agard, D. A. 1987.

Serine protease mechanism: structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. *Biochemistry* 27: 7609-7614.

Brady, K. & Abeles, R. H. 1990. Inhibition of chymotrypsinby peptidyl trifluoromethyl ketones: determinants of slow-binding kinetics. *Biochemistry* 29: 7608-7617.

Brant, D. A., Miller, W. G. and Flory, P. J. 1967. *J. Mol. Biol.* 23: 47.

Brayer, G. D., Delbaere, L. T. J., & James, M. N. G. 1979. Molecular structure of the α-lytic protease from Myxobacter 495 at 2.8 Å resolution. *J. Mol. Biol.* 131: 743-775.

Chothia, C. 1984. Priciples that determine the structure of proteins. *Annu. Rev. Biochem.* 53: 537-572

Connolly, M. L. 1983a. *J. Appl. Crystallogr.* 16: 548-558.

Connolly, M. L. 1983b. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221: 709-713.

Connolly, M. L. 1985. Computation of molecular volume. *J. Am. Chem. Soc.* 107: 1118-1124.

Cybulski, S. M. and Scheiner, S. 1989. Factors contributing to distribution energies of bent hydrogen bonds. Implications for proton-transfer potentials. *J. Phys. Chem.* 93: 6565-6574.

Dorovska, V. N., Varfolomeyev, S. D., Kazanskaya, N. F., Klyosov, A. A. & Martinek, K. 1972. The influence of the geometric properties of the active centre on the specificity of - chymotrypsin catalysis. *FEBS Lett.* 23: 122-124.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., & Wells, J. A. 1986. *Science* 233:

659-663.

Fersht, A. 1985a. *Enzyme Structure and Mechanism* , Freeman, San Francisco, ed. 2.

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., & Winter, G. 1985b. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature (London)* 314: 235-238.

Fersht, A. R. 1988. Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* 27: 1577-1580.

Fersht, A. R., Shindler, J. S. and Tsui, W. -C. 1980. Probing the limits of protein-amino acid side chain recognition with the aminoacyl-tRNA syntetases. *Biochemistry* 19: 5520-5524

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. 1985. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* 183: 479-502.

Hames, B. D., & Rickwood, D., Eds. 1981. *Gel Electrophoresis of Proteins*, IRL, Oxford, England.

Hendrickson, W. A. & Konnert, J. 1981. in *Biomolecular Structure, Function, Conformation, and Evolution* (Srinivasam, R., Ed.) Vol. 1, pp 43-47, Pergamon, Oxford, U.K.

Hunkapiller, M. W., Smallcombe, S. H., Whitaker, D. R., & Richards, J. H. 1973. Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implidcations for the catalytic mechanism of serine proteases. *Biochemistry* 12:

4732-4743.

Hunkapiller, M. W., Forgac, M. D., & Richards, J. H. 1976. Mechanism of action of serine proteases: tetrahedral intermediate and concerted proton transfer. *Biochemistry* 15: 5581-5588.

Jones, T. A., 1982. in *Computational Crystallography* (Sayre, D., Ed.) pp 303-317, Oxford University, Oxford, England.

Kettner, C. A., Bone, R., Agard, D. A., & Bachovchin, W. W. 1988. Kinetic properties of the binding of α-lytic protease to peptide boronic acids. *Biochemistry* 27: 7682-7688.

Kettner, C. A., & Shenvi, A. B. 1984. Inhibition of the serine proteases leukocyte elastase, pancreatic elastase, cathepsin G, and chymotryspin by peptide boronic acids. *J. Biol. Chem.* 259: 15106-15114.

Krieger, M., Chambers, J. L., Cristoph, G. G., & Stroud, R. M. 1974. *Acta Crystallogr., Sect. A: Phys., Diffr., Theor. Gen. Crystallogr.* A30: 740-748.

Leszczynski, J. F. & Rose, G. D. 1986. Loops in globular proteins: a novel category of secondary structure. *Science* 234: 849.

Mitchell, J. B. O. and Price, S. L. 1990. The nature of the N-H...O=C hydrogen bond: an intermolecular pertubation theory study of the formamide/formaldehyde complex. *J. Comp. Chem.* 11: 1217-1233.

Pimentel, G. C. and McClellan, A. L. 1960. *The Hydrogen Bond*, Freeman & Co, London.

Rashin, A. A., Iofin, M. & Honig, B. 1986. Internal cavities and buried waters in globular proteins. *Biochemistry* 25: 3619-3625.

Reed, A. E., Weinhold, F., Curtiss, L. A. and Pochatko, D. J. 1986. Natural bond orbital analysis of molecular interactions. *J. Chem. Phys.* 84: 5687-5703.

Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Ann. Rev. Biophys. Bioeng.* 6: 151-176.

Russell, A. J., Thomas, P. G. & Fersht, A. R. 1987. Electrostatic effects on modification of charged groups in the active site cleft of subtilisin by protein engineering. *J. Mol. Biol.* 193: 803-813.

Schecter, I., & Berger, A. 1967. *Biochem. Biophys. Res. Commun.* 27: 157.

Schulz, G. E. and Schirmer, R. H. 1979. "Principles of Protein Structure" Springer-Verlag, New York.

Silen, J. L., Frank, D., Fujishige, A., Bone, R. & Agard, D. A. 1989. Analysis of prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro region is required for activity. *J. Bacteriol.* 171: 1320-1325.

Smith, J. L., Corfield, P. W. R., Hendrickson, W. A. & Low, B. W. 1988. Refinement at 1.4Å resolution of a model of erabutoxin b: treatment of ordered solvent and discrete disorder. *Acta Crystallogra. Sect. A* 44: 357-368.

Stroud, R. M., Kieg, L. J., & Dickerson, R. E. 1974. The structure of bovine trypsin: electron density maps of theinhibited complex at 5Å and at 2-7Å resolution. *J. Mol. Biol.* 83: 185-208.

Thompson, R. C. 1973. Use of peptide aldehydes to generate transition-state analogs of elastase. *Biochemistry* 12: 47-51.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. 1984. *J. Amer. Chem.*

Soc. 106: 765-784.

Westerik, J. O. & Wolfenden, R. 1972. Aldehydes as inhibitors of papain. *J. Biol. Chem.* 247: 8195-8197.

Whitaker, D. R. 1970. The α -lytic protease of a myxobacterium. *Methods Enzymol.* 19: 599-613.

Williams, J. W. & Morrison, J. F. 1979. The kinetics of tight-binding inhibition. *Methods Enzymol.* 63: 437-467.

Wolfenden, R. 1972. *Acc. Chem. Res.* 5: 10-18.

Wycoff, H. W., Doscher, R., Tsernoglou, D., Inagami, T., Johnson, L., Hardman, K. D., Allewell, N. M., Kelly, D. M., & Richards, F. M. 1967. Design of a diffractometer and flow cell system for X-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S. *J. Mol. Biol.* 27: 563-578.

Zamyatnin, A. A. 1972. Protein volume in solution. *Prog. Biophys. Mol. Biol.* 24: 107-123.

UNIVERSITY OF CALIFORNIA
UC San Francisco LIBRARY

Reference

617332

3 1378 00617 3325

UC San Francisco LIBRARY

UNIVERSITY OF CALIFORNIA