Substrate Selection by the ClpXP Protease: A Tail of Destruction

by

Julia M. Flynn

Sc.B., Biochemistry
Brown University, 1997

SUMBITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AUGUST 2004

Signature of Author: _____
Department of Biology
(30 August 2004)

Certified by: _____
Tania A. Baker
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

Substrate Selection by the ClpXP Protease: A Tail of Destruction

by Julia M. Flynn

Submitted to the Department of Biology on 30 August 2004,
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Biology

ABSTRACT

Intracellular proteolysis plays a vital role in many regulatory pathways, helping cells survive a battery of stresses including oxidative damage, heat shock, and starvation. The majority of cellular proteins are degraded very slowly; however, certain proteins are extremely unstable. The concentration of unstable regulatory proteins can be adjusted quickly in response to altered cellular conditions by changes in their rate of degradation and synthesis. In addition, proteases can remove proteins from the cell when their activities are no longer required. Proteolysis is thus a powerful mechanism to regulate many cellular pathways. To execute these tasks, it is imperative that intracellular proteases select their substrates swiftly and discerningly.

This thesis explores the strategies used by the *Escherichia coli* energy-dependent protease, ClpXP, to correctly select its substrates for destruction. Prior to our work, only a small group of ClpXP substrates were known. To identify a larger group, we captured intact substrates in vivo inside of a $ClpXP^{trap}$. Sequence analysis of these identified substrates combined with peptide binding experiments revealed five common motifs that are directly recognized by ClpXP, representing the first general description of rules governing substrate recognition by this protease.

Direct recognition of these accessible degradation tags can be further modulated by adaptor proteins. SspB is an adaptor protein identified for its ability to enhance the degradation of ssrA-tagged proteins by ClpXP. We dissected the sequence information in the ssrA tag required for recognition by ClpX, SspB, and ClpA, another ClpP partner. The ssrA tag contains contiguous bindings sites for ClpX and SspB, but overlapping sites for ClpA and SspB; this spatial arrangement of signals allows for efficient modulation of proteolysis of ssrA-tagged proteins. Finally, additional substrates whose degradation may be regulated by the adaptor protein SspB were determined by identifying substrates captured in $ClpXP^{trap}$ in an *sspB*$^{+}$ strain but not an *sspB*$^{-}$ strain. This analysis led to the identification of the N-terminal fragment of RseA, the master regulator of the extracytoplasmic stress response, as a protein whose ClpXP-mediated degradation is also enhanced by SspB. Degradation of N-RseA leads to activation of σ^{E} and thus induction of the extra-cytoplasmic stress response.

This thesis work has contributed to the understanding of how intracellular proteolysis is regulated to accommodate the selective degradation of a broad range of substrates. ClpXP uses an assortment of recognition strategies, including degradation tags and adaptor proteins, in a combinatorial fashion to regulate protein degradation.

Thesis supervisor: Tania A. Baker
Title: Professor of Biology

ACKNOWLEDGMENTS

I am extremely grateful to my advisor, Tania Baker, whose energy, patience, and determination has truly been the guiding force behind this project. She has challenged me as a scientist, and has provided critical assistance at every step, be it with experiments, presentations, writing, or life outside the lab. I would also like to thank my co-advisor, Bob Sauer, who has been indispensable throughout my graduate career, meeting with me periodically and providing me with endless scientific advice that have helped shape my project. I thank my committee members, Alan Grossman, Mike Yaffe, and Carol Gross for all their advice and reading of my thesis.

I'd like to thank all the members of the Baker lab, past and present, for being both scientific collaborators and good friends. I'd especially like to thank Samia who has gone through this entire process by my side; her strength has been my source of inspiration. I'd like to thank Igor for being my supreme source of information and bestowing me with a tiny fraction of his protein purification knowledge. And, thanks to all of the Baker lab girls whose spirit, energy, and kindness have made the last four years a joyful experience.

None of this would have been possible without the love and support of my family. My parents have always nurtured my scientific interests and set the standard of dedication and hard work that I strive for. I'd like to thank all my friends who have provided me with laughter, sanity, and the occasional escape from science. Especially my best friend Khama who has supported me and challenged me for the last 12 years.

Finally, I would like to thank my beacon, Dan, whose love and encouragement has been my source of strength to make it through to the end. And my two cats, Everett and Eve, for providing me with their unconditional love, and who, by lying across the keyboard of my computer, have provided me with many much needed breaks from writing my thesis.

TABLE OF CONTENTS

TITLE PAGE 1

ABSTRACT 3

ACKNOWLEDGMENTS 5

TABLE OF CONTENTS 7

LIST OF FIGURES 10

LIST OF TABLES 12

CHAPTER ONE: Introduction 13

Overview 15

Intracellular proteases share a common architecture 17

ClpXP has multiple cellular functions 21

Mechanisms of substrate selection by intracellular proteases 25

Direct recognition of degradation signals by ClpXP. 26
Recognition of C-terminal degradation signals by ClpXP (C-motif 1 & 2). 27
Recognition of N-terminal degradation signals by ClpXP (N-motif 1, 2, & 3). 33
Direct recognition of degradation signals by ClpAP, Lon, HslUV and FtsH. 35
Recognition of C-terminal degradation signals by ClpAP, Lon, HslUV and FtsH. 36
Recognition of N-terminal degradation signals by ClpAP, Lon, HslUV and FtsH. 37
Protease domains that mediate substrate specificity. 40
Regulation of proteolysis by cryptic signals. 42
Regulation of proteolysis by adaptor proteins. 46
Spatial regulation of degradation by ClpXP. 51

Combinatorial recognition of multiple signals 52

σ^S (σ^{38}) 55
σ^H (σ^{32}) 55
σ^E (σ^{24}) 56

Recognition of peptide signals as a common means of regulation 56

Recognition by the proteasome in higher organisms 62

Summary 65

CHAPTER TWO: Overlapping recognition determinants within the ssrA tag allow modulation of proteolysis 67

ABSTRACT 68

INTRODUCTION 69

RESULTS 71

Mutant derivatives of the ssrA tag. 71
SsrA-tag•ClpX recognition. 71
SsrA-tag•ClpA recognition. 72
SsrA-tag•SspB recognition. 75
Requirement for dual recognition of the ssrA tag by SspB and ClpX. 78
SspB inhibits degradation of GFP-ssrA by ClpAP. 81

DISCUSSION 83

Binding determinants for ClpX and ClpA in the ssrA tag. 83
SspB is a bifunctional regulator of substrate recognition. 85
Conservation of ClpX- and SspB-recognition modules within the ssrA tag. 86

ACKNOWLEDGEMENTS 88

EXPERIMENTAL PROCEDURES 89

CHAPTER THREE: Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals 91

ABSTRACT 92

INTRODUCTION 93

RESULTS 96

Protein trapping by ClpXP in vivo. 97
Identification of ClpXP substrates. 98
Many trapped substrates have C-terminal degradation signals. 104
Peptide arrays identify N-terminal ClpX-binding signals. 107
Alignments reveal multiple classes of N-terminal recognition motifs. 107

DISCUSSION 114

Substrate discovery through intracellular trapping. 114
Molecular definition of ClpX-recognition motifs. 115
Trapped proteins and roles for ClpXP-mediated degradation. 119

ACKNOWLEDGMENTS 121

EXPERIMENTAL PROCEDURES 122

CHAPTER FOUR: Modulating substrate choice: The SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation 127

ABSTRACT 128

INTRODUCTION 129

RESULTS 132

SspB influences recognition of a set of ClpXP substrates in vivo. 132
One SspB-dependent substrate is an N-terminal fragment of RseA.. 132
$RseA^{1-108}$ is a substrate for SspB and ClpXP in vitro. 135
SspB and ClpX enhance activation of the σ^E regulon in vivo. 138
SspB forms stable delivery complexes with $RseA^{1-108}$ and with σ^E•$RseA^{1-108}$. 141

DISCUSSION 146

ClpXP and SspB regulate σ^E activity via RseA destruction. 146
Adaptor proteins like SspB expand and regulate the substrate repertoire of proteases. 148

EXPERIMENTAL PROCEDURES 151

ACKNOWLEDGEMENTS 155

APPENDIX 157

I. DEFINING CONSENSUS CLPX-RECOGNITION MOTIFS BY SCREENING PEPTIDE LIBRARIES 157

Screening an ssrA peptide library. 158
Screening of a λO peptide library. 163

Experimental Procedures 167

II. PROBING THE ROLE OF THE α-AMINO GROUP IN RECOGNITION OF N-TERMINAL CLPX–RECOGNITION MOTIFS 169

Abstract 169

Introduction and Results 170

Purification of N-formylated proteins. 170
Purification of fMet-IscS$^{1-11}$–Arc. 171
Degradation of fMet–IscS$^{1-11}$–Arc by ClpXP. 172

Discussion 174

Experimental Procedures 175

III. IDENTIFICATION OF SUBSTRATES THAT MAY DEPEND ON RSSB TO BE TARGETED TO CLPXP FOR DEGRADATION 177

RssB influences recognition of a set of ClpXP substrates in vivo. 177
Uncomplexed L10 and NrdH are in vitro substrates for ClpXP. 179
In vitro degradation of L10 and NrdH does not appear to be stimulated by RssB. 182

Experimental Procedures 184

CHAPTER FIVE: Discussion 187

What is the complete set of sequence rules that governs ClpX-substrate specificity? 188
Where on the surface of ClpX are its substrate binding sites? 189
Are there additional ClpX-adaptor proteins? 189
Are there additional ClpXP substrates under different environmental conditions? 190
Are any of the trapped proteins primarily ClpX-disassembly substrates? 190
What role does ClpXP play in the regulation of specific substrates? 191

REFERENCES 195

LIST OF FIGURES

Figure 1.1.	Structures of ClpX and ClpP.	18
Figure 1.2.	Model of Substrate Degradation by ClpXP.	20
Figure 1.3.	Multiple roles of ClpXP.	22
Figure 1.4.	Scheme for capturing substrates inside $ClpP^{trap}$ in vivo.	24
Figure 1.5.	SsrA tagging system.	29
Figure 1.6.	Determinants in the ssrA-tag recognized by ClpX, SspB, and ClpA.	31
Figure 1.7.	Fraction of trapped ClpXP substrates with N- or C-terminal recognition signals.	35
Figure 1.8.	Certain N-terminal amino acids target proteins for degradation.	38
Figure 1.9.	Modes of substrate recognition by ClpXP.	43
Figure 1.10.	LexA is degraded as fragments.	45
Figure 1.11.	ClpXP adaptor proteins.	48
Figure 1.12.	Regulation of the master cell cycle regulator CtrA in *C. crescentus.*	53
Figure 1.13.	Model of the localization and processing of a pro-protein.	58
Figure 1.14.	SH2 domains recognize short peptide motifs containing a pTyr residue.	60
Figure 2.1.	Degradation of GFP-ssrA variants by ClpXP and ClpAP.	73
Figure 2.2.	Effects of ssrA-peptide mutations on SspB recognition.	76
Figure 2.3.	Degradation of GFP-ssrA (A11D) GFP-ssrA (L9A) in the presence of SspB.	79
Figure 2.4.	SspB inhibits degradation by ClpAP.	82
Figure 2.5.	Recognition determinants within the ssrA tag for ClpX, ClpA and SspB.	87
Figure 3.1.	2D-gel analysis of proteins captured by $ClpP^{trap}$.	97
Figure 3.2.	Western blots of trapped proteins.	102

Figure 3.3. C-terminal recognition signals in trapped proteins. 105

Figure 3.4. N-terminal recognition signals. 108

Figure 3.5. Dps has an N-terminal degradation signal. 110

Figure 3.6. ClpXP degradation of Arc-fusion proteins with wild-type or mutant N-terminal recognition signals. 112

Figure 4.1. Proteins captured by $ClpXP^{trap}$. 133

Figure 4.2. Sequence analysis of the $ClpXP^{trap}$-captured RseA fragments. 134

Figure 4.3. Degradation of purified $RseA^{1-108}$ by ClpXP. 137

Figure 4.4. Model of activation of σ^E mediated through an RseA proteolysis cascade. 139

Figure 4.5. Induction of the σ^E regulon is attenuated in *sspB*$^-$ and *clpX*$^-$ strains. 140

Figure 4.6. SspB forms stable complexes with $RseA^{1-108}$ and $RseA^{1-108}\bullet\sigma^E$. 142

Figure 4.7. SspB interacts with residues 77-99 of RseA. 143

Figure A.1. Certain members of a ssrA library are degraded by ClpXP. 160

Figure A.2. Analysis of randomized λO library. 165

Figure A.3. Processing of newly synthesized proteins in bacteria. 171

Figure A.4. SDS–PAGE analysis of protein purity. 171

Figure A.5. ClpXP protease assays. 173

Figure A.6. 2-D gels of proteins captured by $ClpP^{trap}$ in *rssB*$^+$ and *rssB*$^-$ strains. 179

Figure A.7. L10 is rapidly degraded by ClpXP in vitro and stabilized when in complex with L7/L12. 180

Figure A.8. Degradation of NrdH is not activated by RssB. 181

Figure A.9. Degradation of L10 is not activated by RssB. 181

LIST OF TABLES

Table 1.1.	Five classes of ClpX-recognition motifs.	27
Table 1.2.	Ability of ClpX variants to interact with different recognition signals.	41
Table 3.1.	$ClpXP^{trap}$ associated proteins.	99
Table A.1.	Calculated molecular weights of Sol-M-ssrA-X library members.	162

CHAPTER ONE: Introduction

Overview

Bacteria have a remarkable capacity to sense and respond to shifts in their environment. They make the most of their compact genome, using complex regulatory networks to rapidly turn on and off gene products. These organisms use a variety of tactics to permit growth and survival in a wide range of conditions, including movement towards scarce nutrients, survival without food for long periods, and formation of biofilms in the face of hostile environments. Their extreme adaptability allows them to survive extremes of pH, temperature, and osmotic pressure.

To be able to mount these sophisticated responses with the use of only about 5000 gene products, a bacterial cell must be able to rapidly adjust its protein levels. The mechanism that leads to synthesis of new mRNA transcripts in response to altered cellular conditions is well understood. Induction of transcriptional activity results in production of proteins required to respond to these changing conditions. However, what is less understood, but may be equally important, is the role of protein degradation in rapidly regulating cell physiology. Intracellular protein levels can be efficiently modified in response to changing physiological states by balancing new synthesis with degradation. Following a decrease in the rate of synthesis of a protein, the cellular concentration of a protein with a short half-life will change much more rapidly than that of a slowly degraded protein.

Initially, cellular protein turnover was studied as a mechanism to recycle proteins and replenish supplies of amino acids, especially under conditions of starvation when levels of degradation were found to increase (for review see Goldberg and St John 1976). Since then, further studies have uncovered many regulatory pathways controlled by proteolysis. One of the first discovered examples of a specific proteolytic event being involved in the regulation of gene expression was the RecA-dependent cleavage of the λcI repressor, which leads to induction of the *E. coli* λ phage (Roberts and Roberts 1975). It was soon thereafter found that the induction of the SOS response in response to DNA damage involved the degradation of

certain key repressor proteins (Little et al. 1980). Since these early discoveries, there have been many advances in understanding the role of energy-dependent proteases in a variety of diverse regulatory systems, including cell-cycle control, DNA damage repair, and the stationary phase stress response (for review see Gottesman 1996; Gottesman 2003).

It is essential that substrate selection by these intracellular proteases be tightly coordinated and above all, highly specific. Uncontrolled protein degradation of proteins would wreak havoc on cellular processes and rapidly lead to devastation of any organism. Degradation signals present in a protein's sequence or covalently added to a protein target substrates to specific proteases. In addition, regulatory proteins can assist in this recognition. In eukaryotes, proteins can be targeted to the 26S proteasome by post-translational addition of polyubiquitin (Hochstrasser 1996; Hershko and Ciechanover 1998; Voges et al. 1999). The protein ubiquitin ligases (E3s) have a key role in substrate selection for the proteasome because they are primarily responsible for choosing proteins for ubiquitination (for review see Hochstrasser 1996; Hershko and Ciechanover 1998). In bacteria, the AAA+ proteases bind directly to short peptide recognition sequences that are most commonly located near the N- or C-terminus of substrate proteins (for review see Gottesman, 2003; Flynn et al. 2003). Thus, the proteases themselves are mainly responsible for the specificity of substrate selection.

Exciting advances have been made in recent years in understanding how AAA+ proteases in bacteria target their substrates for degradation. With these studies, we have begun to be able to address the following questions: How do these proteases achieve high selectivity? Why are certain proteins extremely stable whereas others are short-lived? Why are some proteins stable under one set of conditions and then rapidly degraded in response to certain environmental stimuli? In this introduction, I will examine the strategies used by these proteases to correctly choose their substrates for degradation, focusing mainly on the well-characterized protease, ClpXP.

Intracellular proteases share a common architecture.

Cytoplasmic, energy-dependent proteases share a common architecture and mechanism; they contain at least one ATPase domain that binds and unfolds substrates, and then translocates these substrates into a sequestered proteolytic chamber where they are degraded into small peptide fragments (for review see Schirmer et al. 1996; Lupas et al. 1997). *E. coli* has five ATP-dependent proteases, ClpXP, ClpAP, HslUV, Lon and FtsH, each with discrete substrate preferences (for review see Gottesman 1996). These proteolytic complexes contain an ATPase component and a multimeric protease component that can reside on one polypeptide chain, as is the case for Lon and FtsH, or as two separate compartments, as for ClpXP, ClpAP and HslUV (Chin et al. 1988; Katayama et al. 1988; Gottesman et al. 1993; Tomoyasu et al. 1995; Rohrwild et al. 1996).

One of the well-characterized energy-dependent proteases is ClpXP, composed of a molecular chaperone, ClpX, and a protease, ClpP (Fig. 1.1; Gottesman et al. 1993). ClpX is a member of the AAA+ superfamily (ATPases associated with variety of activities) that includes the Clp/Hsp100 family, the Lon family, and metalloproteases such as FtsH (for review see Lupas and Martin 2002). Members of the Clp/Hsp100 family are responsible for unfolding and remodeling proteins, dismantling multimers, and solubilizing aggregates (Wickner et al. 1994; Levchenko et al. 1995; Wawrzynow et al. 1995). All of the members of this family, including ClpX, ClpA, and HslU, are hexameric, ring-shaped proteins that contain one or two AAA+ ATPase motifs (Neuwald et al. 1999; Ortega et al. 2000; Singh et al. 2000; Wang et al. 2001). The AAA+ ATPase modules are composed of an α/β domain followed by a mostly α-helical domain; ATP binds in a cleft between these two domains. The α/β domain contains the highly conserved Walker A and B motifs responsible for hydrolyzing ATP (Neuwald et al. 1999). These ATPases also contain auxiliary domains that are not shared with other AAA+ family members. For example, ClpX and ClpA both contain non-

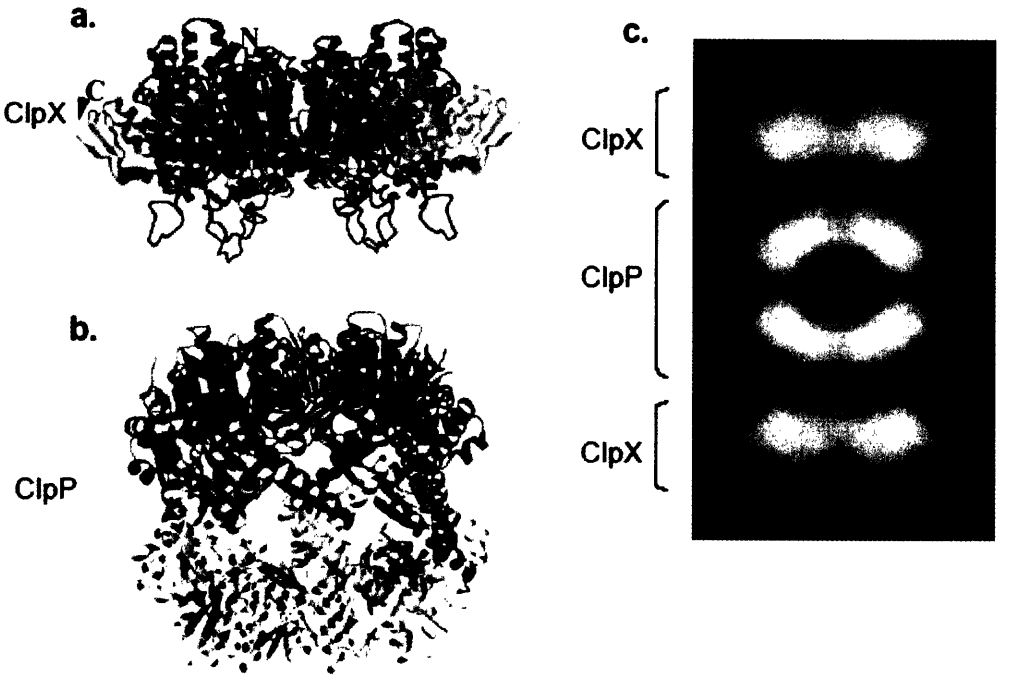


Figure 1.1. Structures of ClpX and ClpP.

(a) A hexameric model of *H. pylori* ClpX viewed from the side. The ATPase core domain, SSD domain, and LGF peptide known to interact with ClpP, are colored in blue, green, and red respectively (from Kim and Kim 2003).

(b) Structure of *E. coli* ClpP viewed from the side. The bottom ring is shown in yellow, while the top ring is colored by subunits (from Porankiewicz et al. 1999).

(c) Electron micrograph of ClpXP. Averaged side-view of ClpXP complexes formed in the presence of ATPγS. ClpX stacks on top of ClpP, so that substrates must first bind ClpX before they can gain access to the proteolytic chamber (from Ortega et al. 2000).

homologous N-terminal domains, whereas HslU has an intermediate domain (I-domain) sandwiched between its two AAA+ modules (Schirmer et al. 1996; Singh et al. 2001). On their own, these ATPases have the ability to act as unfoldases that can bind to and restructure or fully denature substrates. In addition, one or two ATPase components can stack on either side of the proteolytic component to form an active and selective protease complex (Fig. 1.1c) (Grimaud et al. 1998).

The ClpX ATPase assembles on the ClpP protease to form the functional ClpXP degradation machine. ClpP has a barrel-like structure created by two ring-shaped heptamers stacked back-to-back, forming an inner chamber that can accommodate globular proteins as large as 50 kDa (Fig. 1.1b; Maurizi et al. 1990b; Wang et al. 1997). The serine active sites face towards the center of the barrel. The only access to these proteolytic sites in the isolated ClpP molecule is a narrow pore measuring 10 Å across (Wang et al. 1997; Ortega et al. 2000). Thus, only very short peptides are able to diffuse into this pore; even the smallest of folded proteins are not allowed admittance (Thompson and Maurizi 1994). To degrade folded proteins, ClpP must first complex with an ATPase component such as ClpX. ClpX binds directly to the substrate and actively unfolds it and translocates it through the narrow ClpP portal into the sequestered chamber (Fig. 1.2; Gottesman et al. 1993; Wawrzynow et al. 1995; Weber-Ban et al. 1999; Hoskins et al. 2000a).

This sequestration of active sites inside ClpP allows for a high level of regulation of its proteolytic activity, preventing the aberrant degradation of cellular proteins. Because the active sites of the protease subunits are not accessible in their absence, the ATPase components can be thought of as the "gatekeepers" of the proteases. The ATPases are solely responsible for substrate discrimination (for references, see Gottesman 1996). ClpX and ClpA can both form complexes with ClpP and select distinct sets of substrates for degradation by the same proteolytic chamber (for review see Gottesman 1996). HslU chooses substrates for its partner protease, HslV (Missiakas et al. 1996).

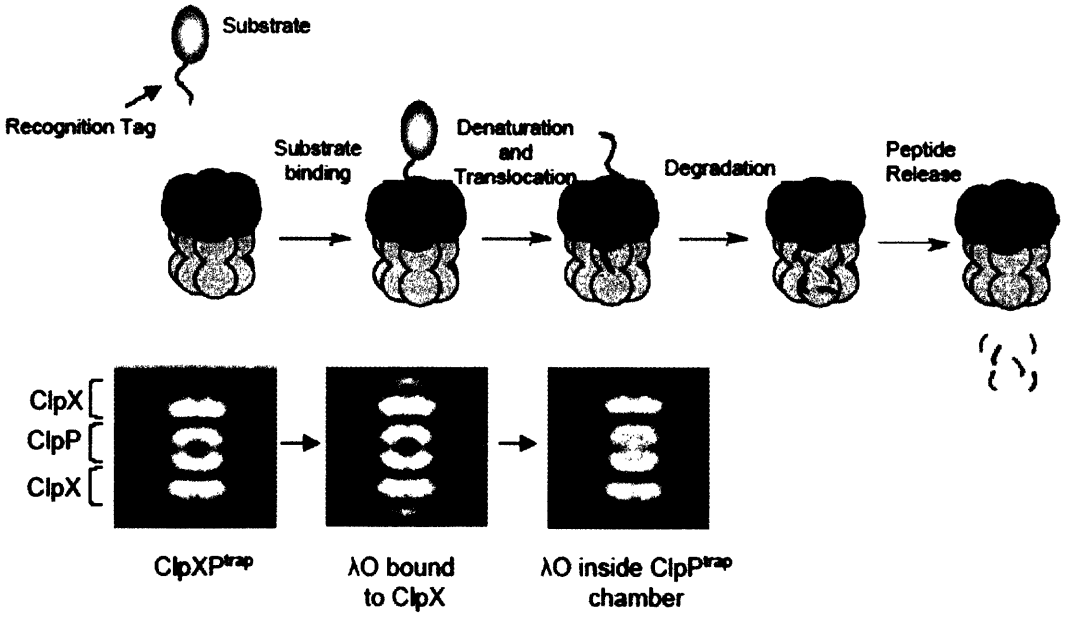


Figure 1.2. Model of Substrate Degradation by ClpXP.

Upper panel: Substrates bind to ClpX, are denatured and translocated into the ClpP chamber, where they are hydrolyzed and released as small peptides.

Lower panel: Electron micrographs showing translocation of substrates into $ClpXP^{trap}$. $ClpXP^{trap}$ was assembled in the presence of ATPγS (left panel). λO was added and images were obtained after 0 min (center panel) and 20 min (right panel). At 0 min, λO can be seen bound to ClpX at either end of the complex as indicated with arrows. At 20 min, λO has been translocated into the $ClpP^{trap}$ chamber (images from Ortega et al. 2000).

ClpXP has multiple cellular functions.

ClpX and *clpP* are not essential genes in *E. coli*, however, *clpP*-defective cells display a number of stress-related phenotypes, including delayed recovery from stationary phase and following a shift to nutrient poor media, a defective ability to form biofilms, and an enhanced sensitivity to UV irradiation (Damerau and St John 1993; Neher et al. 2003a; R. Burton, unpublished data). ClpP is also important for the virulence of a number of bacterial pathogens, is required for cell-cycle progression in *Caulobacter crescentus*, and plays a role in development in *Bacillus subtilis* (for review, see Porankiewicz et al. 1999; also see Jenal and Fuchs 1998; Msadek et al. 1998).

Despite these diverse phenotypes, until recently, only a small handful of ClpXP substrates had been identified. However, examination of these substrates hinted at important roles for ClpXP in a diverse array of cellular processes. ClpXP was originally discovered as a component required for ClpP-dependent degradation of the λO phage replication protein. In vitro degradation of λO was used as a biochemical assay to purify the enzyme responsible for this activity from cell lysate (Gottesman et al. 1993). λO has a half-life as short as one to two minutes in wild-type *E. coli* cells, while in *clpX* and *clpP* mutant cells this replication initiation protein is stable for over an hour (Wyatt and Inokuchi 1974; Wegrzyn et al. 1992; Wojtkowiak et al. 1993).

Following this initial characterization of ClpXP, four additional phage or plasmid-encoded proteins (Mu repressor, Mu transposase (MuA), RK2 replication protein TfrA, and the P1 antidote protein PhD) and three *E. coli* proteins (the stationary phase sigma factor σ^S, the SOS protein UmuD' and a type I restriction-modification subunit HsdR) were identified as ClpXP substrates (see Gottesman 1996) and references therein; (Frank et al. 1996; Konieczny and Helinski 1997; Makovets et al. 1998). In addition, ClpXP was found to be responsible for the degradation of ssrA-tagged proteins (Gottesman et al. 1998). These few

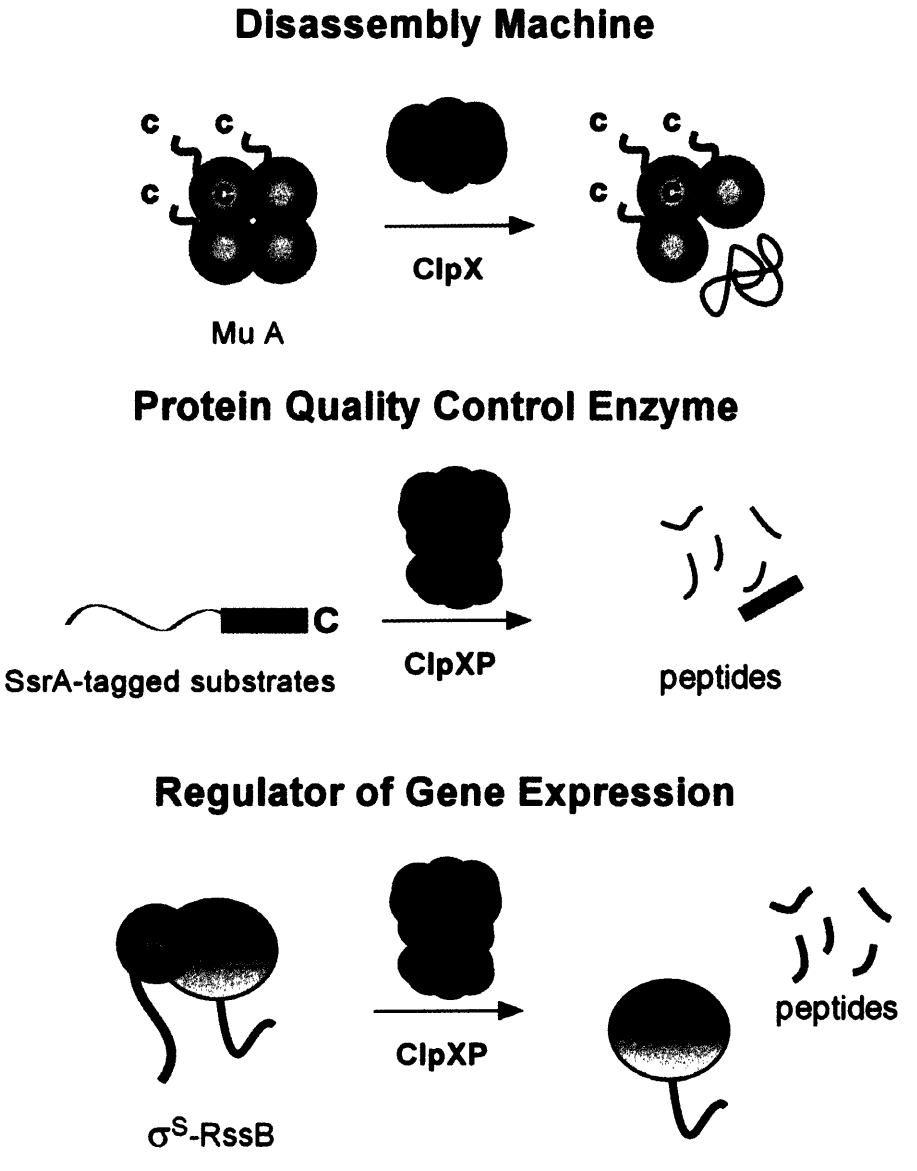


Figure 1.3. Multiple roles of ClpXP.

identified substrates began to reveal the multiple functions of ClpX in the cell: ClpX as a disassembly chaperone (MuA), as a regulator of gene expression (σ^S), and as a protein quality control enzyme (ssrA) (Fig. 1.3; (Levchenko et al. 1995; Schweder et al. 1996; Gottesman et al. 1998).

MuA transposase is a monomeric protein that assembles into a tetramer upon binding the ends of the Mu genome, and catalyzes the transfer of the ends of the phage's DNA into a new DNA site (Craigie and Mizuuchi 1987; Surette et al. 1987). Once this recombination reaction is complete, ClpX disassembles this hyper-stable protein-DNA complex, allowing phage DNA replication to begin (Levchenko et al. 1995). This restructuring activity of ClpX does not require ClpP, although ClpXP is able to degrade MuA monomers in vitro. Thus, ClpX can function alone as a disassembly machine, or together with ClpP as part of a protease. A well-characterized ClpXP degradation substrate is the stationary phase sigma factor, σ^S. σ^S is rapidly degraded by ClpXP during exponential growth conditions, and is greatly stabilized during stationary phase when its activities are required (Schweder et al. 1996). In this case, the proteolytic activity of ClpXP is performing a regulatory role. Degradation by this protease can also provide a more general protein quality control function in the cell. ClpXP is the main protease responsible for degrading proteins marked for destruction by the ssrA tag, a natural in vivo tagging system (see below for a more detailed description) (Gottesman et al. 1998).

Our knowledge of the diverse roles ClpXP plays in controlling cellular processes was greatly enhanced by the identification of a larger group of substrates using an in vivo trapping procedure presented in this dissertation (see Chapter Three). An inactive form of ClpP ($ClpP^{trap}$) was successfully used to capture ClpXP substrates in vivo and thus quickly identify many ClpXP substrates (Fig. 1.4). 2-D gels of proteins captured by $ClpP^{trap}$ under different conditions provided a snapshot of the ClpXP substrates degraded during these conditions. The proteins trapped under "normal" growth conditions included transcription factors,

metabolic enzymes, and proteins involved in the starvation and oxidative stress responses. For example, a set of ClpXP substrates trapped under these conditions are proteins normally active during stationary phase. One of these substrates, Dps, is a DNA binding protein that protects DNA against many environmental stresses such as oxidative damage. Transcription of Dps by σ^S leads to greatly enhanced levels of this protein during stationary phase growth (Almiron et al. 1992). As cells recover from stationary phase and re-enter logarithmic growth, Dps is rapidly degraded by ClpXP (see Chapter Three). Here, degradation is playing an important role in re-adjusting the levels of Dps upon alteration of cellular conditions.

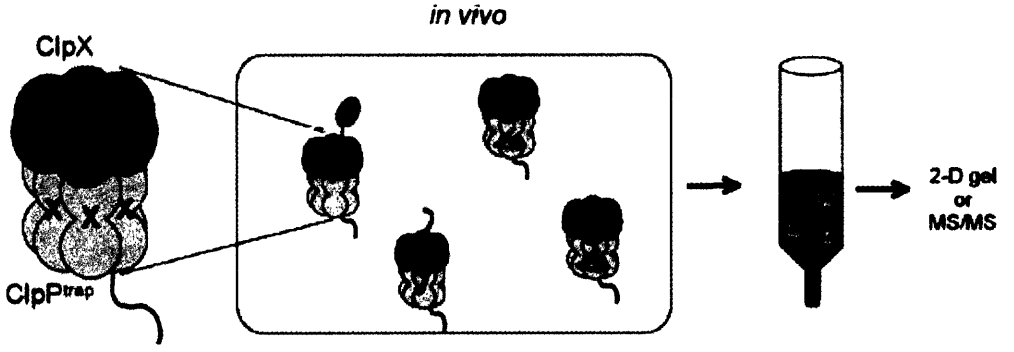


Figure 1.4. Scheme for capturing substrates inside ClpP^{trap} in vivo.

An inactive and epitope-tagged form of ClpP was expressed in vivo to capture substrates. ClpP^{trap} was then purified from the cells and trapped proteins were identified by tandem mass spectrometry (see Chapter Three).

These trapping experiments indicate that a number of ClpXP substrates are proteins whose cellular levels are induced during various stresses. It is likely that turnover by ClpXP keeps levels of these proteins low in non-stress conditions. Altering the rate of degradation of these proteins along with their rate of synthesis can allow for rapid responses coupled to changes in cellular conditions. One important role of ClpXP is thus to change the proteome in response to a variety of stresses. Capturing proteins under diverse stress conditions will likely increase the repertoire of identified ClpXP substrates.

Mechanisms of substrate selection by intracellular proteases.

Due to the destructive nature of proteolysis, many mechanisms must be in place to ensure that the intracellular degradation machinery is, above all, highly selective. The fact that intracellular proteases reside in the same compartment as their substrates requires that there is a high degree of regulation of their proteolytic activity. Conventional proteases such as trypsin are not well-suited for this activity; these proteases cleave following certain amino acids in exposed regions of all proteins and thus would non-specifically destroy all proteins, obliterating the host cell. Instead, intracellular proteases such as ClpXP specifically choose their target substrates and processively degrade them so the substrate is completely destroyed. To ensure this specificity, proteases must have recognition mechanisms in place to readily distinguish a substrate from a non-substrate.

Intracellular proteases in bacteria normally interact with sequences in substrates known as "recognition signals" or "degradation tags." These are intrinsic peptide sequences that have been shown through genetic analysis to be necessary for the degradation of the protein. A true degradation tag is also sufficient to target an otherwise stable protein for proteolysis.

Despite this specificity of signal recognition, intracellular proteases must maintain the ability to degrade a broad range of substrates. The large number of substrates captured by

ClpXP^{trap} illustrates the diversity of proteins with which ClpXP must interact (see Chapter Three). In addition, recognition of these accessible degradation tags must be coordinated with environmental cues. In the following sections, I will discuss the degradation signals directly recognized by ClpXP and other intracellular proteases, followed by mechanisms such as adaptor proteins and cryptic signals these proteases use to appropriately target substrates for destruction.

Direct recognition of degradation signals by ClpXP.

Of the five intracellular ATP-dependent proteases in *E. coli*, the substrate specificity of ClpXP is the most extensively studied. From what is so far understood, ClpXP also appears to be the most selective of these proteases. The known ClpXP substrates are principally native proteins and thus, ClpXP must recognize "destruction" signals in the folded protein. In fact, thermodynamic stability has little effect on the susceptibility of a substrate to proteolysis by ClpXP, a fact that has implications not only on the mechanism of degradation by ClpXP, but also suggests that a substrate need not be unfolded to be recognized (Burton et al. 2001).

ClpXP interacts with a diverse set of signals. The ClpX-recognition tags primarily range from 3-10 amino acids in length and are most often positioned near the extreme N- or C-terminus of a protein. This precise locale derives from two sources: 1) These are often the most accessible regions of a protein and are least likely to be buried within a native protein and 2) The α-carboxyl and α-amino groups found only at the N- and C-terminus of a protein could in principle provide unique molecular determinants for substrate recognition. The known ClpXP-recognition signals have been divided into five classes of sequences, two located at the C-terminus of substrates, C-motif 1 and 2, and three located near the N-terminus, N-motif 1, 2 and 3 (Table 1.1). The following sections will discuss the advances made in recent years in the characterization of the primary degradation signals recognized by ClpXP and other proteases.

Table 1.1. Five classes of ClpX-recognition motifs.

Motif	Consensus	Model Substrates	Sequence
N-motif 1	Polar-T/Φ-Φ-+-Φ	λO Dps	NTAKI STAKL
N-motif 2	Met-+-Φ-Φ-X_5-Φ	OmpA IscS	NH_2-MKKTAX_5V NH_2-MKLPIX_5A
N-motif 3	Φ-X-Polar-X-Polar-X-+-Polar	DksA	NH_2-MQEGQNRK
C-motif 1	Φ-Φ-Φ-COOH	ssrA N-RseA	LAA-COOH VAA-COOH
C-motif 2	+-+-+-+-Φ-Φ	MuA YbaQ	RRKKAI-COOH RAKKVA-COOH

+ = basic amino acid
Φ = hydrophobic amino acid
X = any amino acid

Recognition of C-terminal degradation signals by ClpXP (C-motif 1 & 2).

The C-motif 1 class of signals is based on similarity to the known ClpX-recognition signal in the ssrA tag. This motif is defined by two to three C-terminal nonpolar amino acids; small, uncharged residues with a predominance of alanines occupy the two C-terminal residues while hydrophobic residues such leucine, are more common at the third residue from the C-terminus (see Chapters Two & Three and Appendix I for a more complete analysis). The C-motif 2 class is defined by similarity to the ClpX-recognition signal at the C-terminus of MuA. Proteins with this signal have nonpolar C-terminal dipeptides and basic side chains in the region three to six residues before the C-terminus (see Chapter Three; Table 1.1).

The observation that the C-terminal sequence of a protein can influence its susceptibility to proteolytic activity first occurred in the late 1980's. Fusing a 25 amino acid

C-terminal "tail" to a protein with a short in vivo half-life greatly stabilized the protein against degradation (Parsell and Sauer 1989). Parsell *et al.* performed a random mutagenesis analysis of the C-terminal residues of this tail and measured the in vivo turnover of these proteins (Parsell et al. 1990). It was mainly the polar character of these residues that influenced the protein's sensitivity to proteases. A fusion protein with a C-terminal sequence, Trp-Val-Ala-Ala-Ala was rapidly degraded, whereas the half-life of the protein with the original stabilizing tail (Arg-Ser-Glu-Tyr-Glu) was greater than 600 minutes.

The positioning of the nonpolar sequence at the very C-terminus was critical to its destabilizing effect. Charged amino acids such as aspartate had the most stabilizing affect at the very C-terminal residue, and this affect gradually diminished in relation to its distance from the C-terminus. The most destabilizing residues, Ala, Cys and Val, were not the most hydrophobic amino acids, indicating no direct correlation between hydrophobicity and intracellular stability. Thus, it was found, that small, nonpolar amino acids, positioned at the extreme C-terminus of a protein, can target that protein for rapid degradation.

The most well-characterized C-terminal ClpX-recognition signal is that of the ssrA tag, a natural tagging system in which an otherwise stable protein can be destabilized by addition of a C-terminal sequence. The ssrA tag is an 11 amino acid peptide, AANDENYALAA, added co-translationally onto polypeptides stalled on the ribosome during translation (Keiler and Sauer 1996). SsrA RNA has a dual nature as both a tRNA molecule that is chargeable with alanine, and an mRNA molecule that codes for the last 10 amino acids of the ssrA tag. When a ribosome stalls during translation, due to, for example, an incomplete message or a rare codon, aminoacylated-ssrA RNA is recruited to the ribosome and the nascent chain is transferred onto the alanine-charged tRNA. Translation then switches to the reading frame in the ssrA RNA. Thus, the ssrA-tagging system clears stalled mRNAs off the ribosomes allowing translation to resume (Fig. 1.5; for review see Karzai et al. 2000).

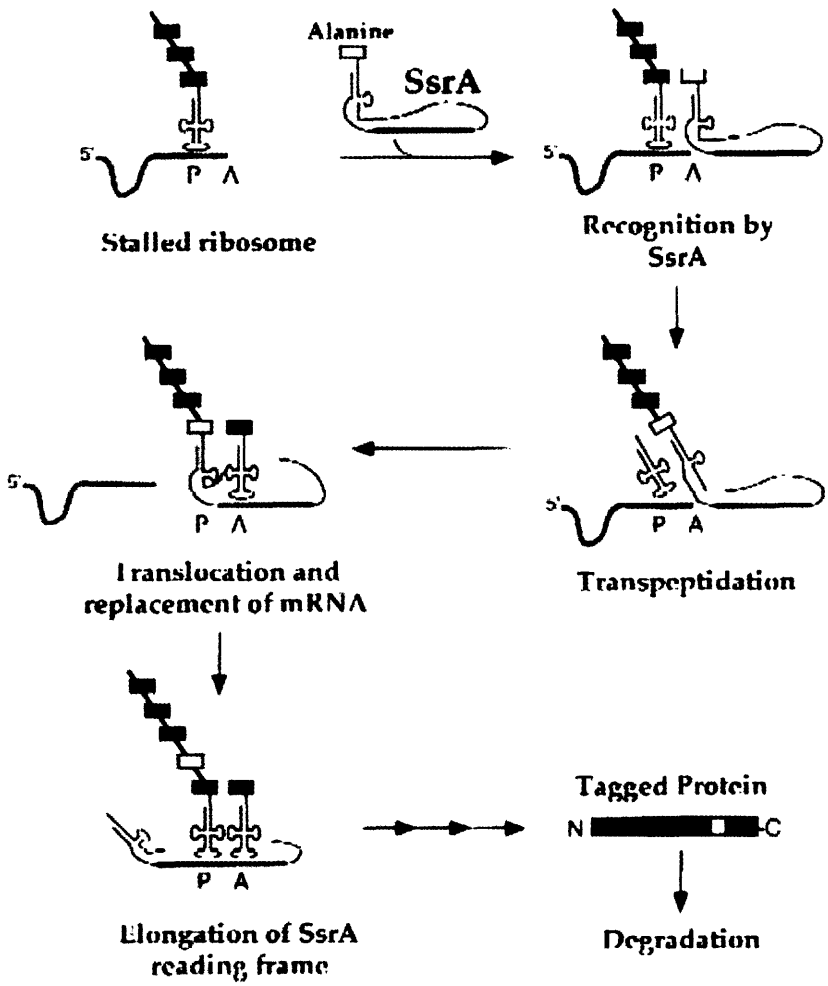


Figure 1.5. SsrA tagging system.

When a ribosome stalls during translation aminoacylated-ssrA RNA is recruited to the ribosome and the nascent chain is transferred onto the alanine-charged tRNA. Translation then switches to the reading frame in the ssrA RNA. The resulting ssrA-tagged protein is targeted for degradation (figure from Karzai et al. 2000).

Shortly after this tagging system was identified, it was observed that the C-terminal sequence of the ssrA tag, Tyr-Ala-Leu-Ala-Ala, was similar to the C-terminal sequences found by Parsell *et al.* to rapidly target a protein for degradation (Parsell et al. 1990; Keiler et al. 1996). Accordingly, an ssrA-tagged protein was found to be degraded in vivo with a half-life of less than five minutes, whereas a control tag terminating in Asp-Asp was stable for over an hour (Keiler et al. 1996). Thus, a second role of the ssrA-tagging system is to target possibly deleterious polypeptide fragments for degradation.

The bulk of degradation of ssrA-tagged proteins in vivo is achieved by ClpXP and ClpAP (Gottesman et al. 1998). Even in Clp-deficient strains, a small residual level of degradation of ssrA-tagged proteins was observed. This remaining proteolytic activity is likely due to degradation by FtsH and Tsp, respectively membrane and periplasmic proteases that have the capability of degrading ssrA-tagged proteins (Keiler and Sauer 1996; Herman et al. 1998).

Purified ClpXP and ClpAP complexes are both capable of degrading ssrA-tagged proteins in vitro (Gottesman et al. 1998). In fact, the ssrA tag is a strong primary recognition signal; all the information required for degradation of ssrA-tagged proteins is encoded within the ssrA tag itself and is independent of factors such as the stability of the attached protein or other recognition signals within the protein. Interaction of this tag with ClpXP or ClpAP can thus result in the unfolding and degradation of any attached protein including highly stable proteins such as Green Fluorescent Protein (GFP) (Gottesman et al. 1998; Weber-Ban et al. 1999; Kim et al. 2000; Burton et al. 2001). Interestingly, although ClpXP and ClpAP degrade ssrA-tagged proteins at a similar rate in vitro, these substrates are preferentially degraded by ClpXP in vivo (Gottesman et al. 1998; Flynn et al. 2001). This phenomenon will be discussed further below in the section on adaptor proteins.

Further dissection of the ssrA tag by mutational analysis, as will be discussed in Chapter Three, revealed that the ClpX-binding determinants of the ssrA tag are highly

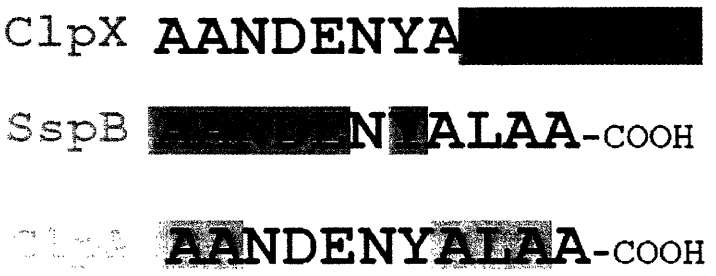


Figure 1.6. Determinants in the ssrA tag recognized by ClpX, SspB, and ClpA.

Residues determined to be important for recognition of the ssrA tag by mutational analysis by ClpX, SspB, or ClpA, are highlighted in purple, blue, and green respectively. SspB and ClpX bind to adjacent sequences in the tag, working together to enhance degradation of tagged proteins by ClpXP. SspB masks the ClpA-recognition determinants, inhibiting degradation of ssrA-tagged proteins by ClpAP (see Chapter Two).

localized to the extreme C-terminus of the tag, Leu^{9}-Ala^{10}-Ala^{11}-COOH (Fig. 1.6). Mutation of either of the C-terminal alanines to aspartates completely obliterates recognition of the tag by ClpX.

A role of C-terminal nonpolar residues in recognition by ClpXP was beginning to emerge. About half of the known ClpXP substrates shared similar nonpolar side chains at the penultimate and C-terminal residues. The bacteriophage Mu repressor has a C-terminal sequence of KKAV-COOH, and is degraded by ClpXP (Laachouch et al. 1996). In addition, a mutant form of the Mu repressor with a C-terminal sequence of RKVL-COOH, resulting from a frameshift mutation near the 3′ end of the gene, is rapidly degraded by ClpXP. Fusing this

C-terminal sequence onto an otherwise stable reporter protein, confers its sensitivity to ClpXP (Laachouch et al. 1996).

The C-terminus of MuA also contains a ClpX-recognition signal. The C-terminal amino acids of MuA are RRKKAI-COOH. Deletion of the last four amino acids of this sequence rendered MuA refractory to both disassembly by ClpX and degradation by ClpXP (Levchenko et al. 1995). Further mutational dissection of this tag indicated that both the nonpolar C-terminal residues and the stretch of basic residues are important determinants for ClpX's recognition of this tag (Levchenko et al. 1997b). The Mu repressor substrates also bear a similar stretch of basic residues. Thus, these recognition signals contain both similar and distinct sequences compared to the ssrA tag, indicating that these C-terminal sequences could form a distinct category of ClpX-recognition signals.

Analyzing the important residues for recognition within this small group of known ClpXP substrates lay the foundation for understanding the molecular determinants that characterize substrate selectivity by ClpX. However, further definition of these signals required analysis of a larger sample population. Identification of many new substrates, as presented in Chapter Three of this thesis, has made this analysis possible. Inspection of the C-termini of the trapped proteins indicated that 50% of these proteins contain C-terminal recognition tags that fall into two distinct classes: C-motif 1, defined by homology to the recognition determinants in the ssrA tag, and C-motif 2, which is more similar to the C-terminal tail of MuA (Table 1.1). Representative members of these two motif classes were found to be sufficient to target a stable reporter protein for degradation by ClpXP. Further characterization of these motif classes is presented in Chapter Three and Appendix I.

ClpXP uses a similar mode of signal recognition in other species of bacteria. The C-terminal three amino acids of the ssrA tag are highly conserved among a variety of bacterial species that all contain ClpX orthologs (LAA or VAA) (Karzai et al. 2000). It is likely that ClpX recognizes these residues similarly in all bacteria. Accordingly, in *B. subtilis*, both deletion of

clpX or mutation of the C-terminal alanines to aspartates, stabilizes ssrA-tagged proteins (Wiegert and Schumann 2001).

A number of *Caulobacter crescentus* proteins contain C-motif 1 signals that are recognized by ClpXP. ClpXP in *C. crescentus* is required for cell cycle progression and viability (Jenal and Fuchs 1998). ClpXP substrates in *C. crescentus* include the essential cell cycle regulator, CtrA, and the chemotaxis receptor coupling protein and response regulator, CheW and CheY (Domian et al. 1997; Alley 2002). Each of these substrates have the C-terminal non-polar amino acids VAA or LAA that signal their degradation by ClpXP. Mutating the C-terminal Ala-Ala residues of CtrA to Asp-Asp stabilizes the transcription factor throughout the cell cycle (Domian et al. 1997).

Recognition of N-terminal degradation signals by ClpXP (N-motif 1, 2, & 3).

We defined three new classes of ClpX-recognition signals, N-motif 1, N-motif 2, and N-motif 3, based mainly on the experiments described in Chapter Three. The defining characteristics of these motifs and representative proteins containing these degradation signals are depicted in Table 1.1. These N-terminal motifs are based on alignment, peptide binding, sufficiency for degradation, and mutational analysis. Further mutational analysis such as that presented in Appendix I is needed to fully characterize the amino acid requirements at each position.

Initial hints into the complexity of substrate recognition by ClpXP came from studies performed on the first identified ClpXP substrate, λO (Wojtkowiak et al. 1993). This was the only previously identified substrate that appeared to have a ClpX-recognition signal near its N-terminus. Deletion of the first 18 amino acids of λO stabilized it against hydrolysis by ClpXP. In addition, the N-terminal portion of λO was degraded much more efficiently by ClpXP than the C-terminal region (Gonciarz-Swiatek et al. 1999). These studies alluded to the existence of an N-terminal ClpX-recognition signal.

A more in-depth understanding of N-terminal recognition signals was greatly facilitated by our studies identifying a larger group of ClpXP substrates (see Chapter Three). In a peptide filter binding assay, ClpX bound the N-terminal peptides of about 60% of the substrates identified in the trap, suggesting that many ClpXP substrates are recognized through N-terminal sequences (Fig. 1.7). Alignment of the sequences that bound ClpX revealed the motifs discussed above: N-motif 1, 2, and 3. Representative sequences from each motif were found to be sufficient for targeting a stable reporter protein for degradation, demonstrating that these sequences are functional ClpX-recognition signals. For instance, the 11 N-terminal residues of λO (an N-motif 1 sequence) converted a reporter protein into a ClpX substrate, supporting the deletion analysis of λO that designated a role for this sequence in ClpX recognition. There is no evidence for a direct role of the N-terminus in recognition of these N-motifs, indicating that it is more likely that the location of the N-terminal signals is due to the increased accessibility of this region (see Appendix II).

Our analysis has shown that greater than 90% of ClpXP substrates have peptide sequences near their N- or C-termini that target them for degradation and about 25% have degradation signals at both termini (Fig. 1.7). For substrates that carry two signals, it is possible that one signal may be the primary sequence that engages the substrate for unfolding and degradation, while the other simply tethers the substrate to the protease, increasing the binding affinity. This would be analogous to a case in which a substrate with a primary signal is tethered to ClpX by an adaptor protein, as will be discussed below. An emerging theme in substrate recognition by ClpXP appears to be the combinatorial recognition of multiple weak degradation signals.

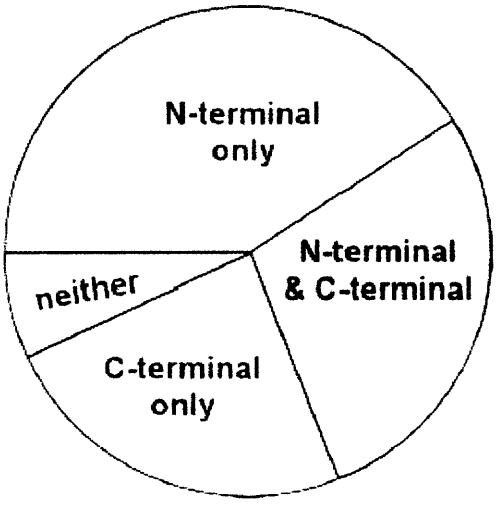


Figure 1.7. Fraction of trapped ClpXP substrates with N- or C-terminal recognition signals.

Direct recognition of degradation signals by ClpAP, Lon, HslUV and FtsH.

Much less is understood of the requirements for substrate recognition by ClpAP, Lon, HslUV and FtsH. A well-defined primary recognition motif for each of these proteases has yet to emerge. However, the same general strategies for recognition appear to be used for all of the bacterial intracellular proteases. When localized, the recognition motifs in these substrates appear to be short peptide sequences near the N- or C-terminus of the substrate. There is a significant level of redundancy between these proteases; many unstable proteins in *E. coli*, including ssrA-tagged proteins (see Chapter Two), RseA (see Chapter Four) and SulA, are degraded by multiple proteases. However, each of these proteases also has the ability to recognize distinct substrates; for instance, ClpA does not recognize C-motif 1 in the ssrA tag, or C-motif 2 in MuA (Flynn et al. 2001 and I. Levchenko, unpublished data). Thus, current data suggests that the precise peptide motifs that target proteins for degradation by ClpXP and ClpAP are different, signifying that these proteases must contain substrate-binding pockets with different recognition characteristics. Unlike ClpXP, Lon and ClpAP both have the

ability to recognize and degrade unfolded proteins, and furthermore, Lon is known to degrade damaged proteins (see Gottesman 1996 for review; also see Charette et al. 1981; Katayama et al. 1988; Hoskins et al. 2000b).

Recognition of C-terminal degradation signals by ClpAP, Lon, HslUV and FtsH.

Although ClpXP is the main protease that degrades ssrA-tagged proteins in vivo, ClpAP and FtsH can efficiently degrade these proteins in vitro (Gottesman et al. 1998; Herman et al. 1998). As reported in detail above, mutational analysis showed that ClpX recognizes the three C-terminal residues of the ssrA tag. In contrast, ClpA interacts with residues 1-2 at the N-terminus and 9-11 at the C-terminus of the tag (see Chapter Two; Fig. 1.6). Thus, although both ATPases interact with a common tag, they do so by interacting with distinct sequences in the peptide. This indicates that ClpA and ClpX may have distinct substrate binding pockets that allow them to interact with distinct groups of substrates. This analysis also indicates that ClpA may not require the free α-carboxylate as a recognition determinant of the ssrA tag, putting ClpX at a unique position over ClpA to degrade proteins first cleaved by other proteases, a recognition mechanism discussed below.

It appears that the sequence requirements for FtsH are much more relaxed than for ClpXP. A detailed analysis of sequence requirements of the ssrA tag for FtsH has not been performed. However, a group of non-polar tails fused to the C-terminus of a reporter protein were studied for degradation by FtsH (Herman et al. 1998). FtsH was able to degrade all the proteins tested, while ClpXP was only able to degrade a subset of them. Thus, the sequence selectivity of FtsH appears to be lower than that of ClpXP. FtsH lacks the ability to robustly unfold proteins; it is not able to degrade the thermodynamically stable GFP-ssrA, while it efficiently degrades the less stable λcI-ssrA (Herman et al. 2003). This finding indicates that FtsH may select its substrates primarily based on thermodynamic stability, in contrast to ClpXP, which selects its substrates based on sequence and not stability.

SulA is an inhibitor of cell division; its levels are closely coupled to DNA damage by a combination of transcriptional regulation and degradation by Lon and HslUV proteases (Mizusawa and Gottesman 1983; Huisman et al. 1984; Wu et al. 1999). Mutating the C-terminal histidine of SulA stabilizes the substrate against degradation by Lon. In addition, L-histidine is a competitive inhibitor of this degradation. However, a C-terminal histidine residue is not sufficient to target any protein to Lon, implying that once the histidine is recognized by Lon, an additional determinant must be involved in regulating its degradation (Ishii and Amano 2001). Although this data hints at an interesting C-terminal Lon-recognition signal distinct from that of the other proteases, it is clear that there is much to be learned regarding substrate recognition by this protease. The ability of these proteases to interact with different signals is integral to their biological roles which require them to select distinct groups of substrates for degradation.

Recognition of N-terminal degradation signals by ClpAP, Lon, HslUV and FtsH.

By fusing various amino acids to the N-terminus of β-galactosidase, Vashavsky *et al.* identified a proteolysis recognition mechanism known as the "N-end rule" that is present in all organisms from bacteria to mammals (Tobias et al. 1991; Varshavsky 1992). This rule relates the in vivo half-life of a protein to the identity of its N-terminal amino acid. In *E. coli*, the destabilizing N-terminal amino acids, or "N-degrons," are the bulky hydrophobic residues Leu, Phe, Trp, and Tyr. In addition, an enzyme known as L/F transferase can conjugate Leu or Phe onto proteins bearing an N-terminal Arg or Lys to enhance their instability (Fig. 1.8). ClpAP appears to be the protease that degrades these substrates carrying N-degrons. Knocking out *clpA* or *clpP* inactivates this pathway (Tobias et al. 1991). However, to date, there are no known physiological substrates of this pathway.

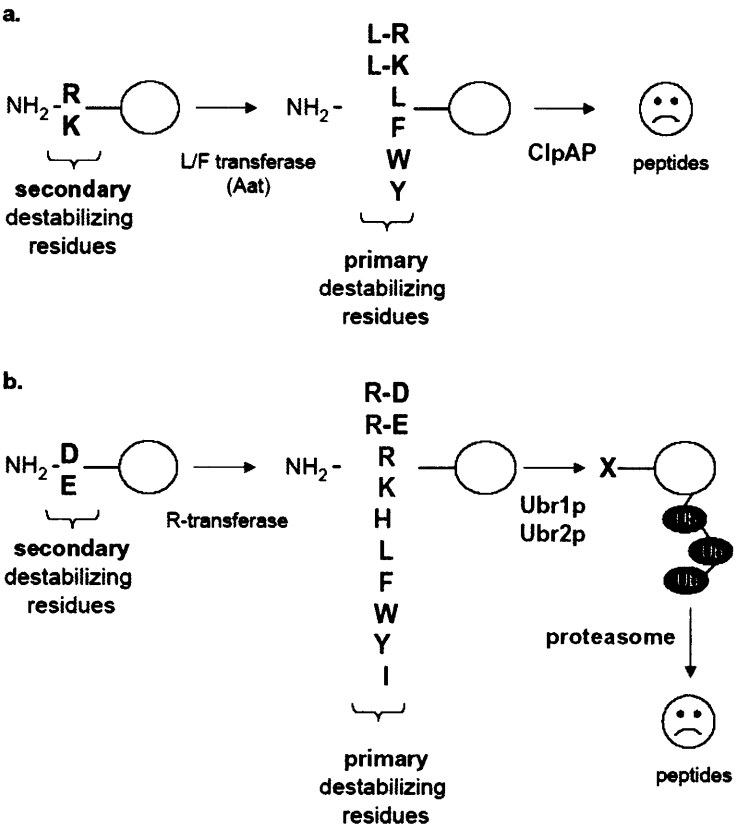


Figure 1.8. Certain N-terminal amino acids target proteins for degradation.

(a) The *E. coli* N-end rule pathway. Substrates bearing primary destabilizing residues are degraded by ClpAP.

(b) The *S. cerevisiae* N-end rule pathway. Substrates bearing primary destabilizing residues are bound by the E3 ligase Ubr1p, ubiquitinated, and targeted to the proteasome. Ubr1 has three substrate binding sites; one that binds substrates with basic N-terminal residues, one for hydrophobic N-terminal residues, and one for non N-end rule substrates (adapted from Varshavsky 1996).

Other ClpAP substrates have recognition signals near the N-terminus dissimilar to N-end rule signals. The plasmid P1 initiator protein, RepA, is an in vitro degradation substrate of ClpAP. In addition, ClpA alone is able to dissociate inactive RepA dimers into active monomers (Wickner et al. 1994). The first 15 amino acids of RepA are both necessary for its degradation by ClpAP and are sufficient to target a reporter protein to ClpAP for degradation (Hoskins et al. 2000a). The first 18 amino acids of HemA (Glutamyl-tRNA reductase) also appear to contain a recognition signal that targets the protein to degradation by ClpAP (Wang et al. 1999). However, there are no clear similarities between these two recognition sequences or with the N-terminal ClpX-recognition sequences. Much more work is needed to understand specific substrate recognition by this protease.

FtsH is a membrane-bound protease whose active sites face the cytoplasmic face of the membrane (Tomoyasu et al. 1995). It has both membrane and cytosolic substrates, and is the only essential energy-dependent protease in *E. coli* (Herman et al. 1995; Kihara et al. 1995; Kihara et al. 1997). To degrade membrane proteins, it appears that FtsH recognizes cytoplasmic tails with lengths greater then 20 amino acids, often located at the N-terminus of the substrate. FtsH can recognize a diverse array of amino acid sequences as long as the tail is longer than 20 residues, indicating that it is the length, and not the exact sequence that is important for this interaction (Chiba et al. 2000). As discussed above, FtsH lacks the ability to degrade stably folded proteins, and thus it has been suggested that this protease can generically recognize proteins with long cytoplasmic tails and then assess their folded state as a secondary decision towards substrate selection (Herman et al. 2003).

Each of the five intracellular *E. coli* proteases are programmed to perform different, yet overlapping, biological roles. The fact that there are five proteases with distinct substrate specificities greatly expands the diversity of primary signals that can be recognized. In addition, their overlapping specificities ensure the complete proteolysis of important substrates. Changing the cellular levels of different proteases in response to different

conditions is an efficient method to change the profile of degradation substrates, effectively modifying the content of the proteome.

Protease domains that mediate substrate specificity.

How do these proteases mediate the direct interaction with such a broad range of recognition signals? A number of studies indicate that the ATPase subunits have at least a couple of distinct substrate binding pockets on their surface (Levchenko et al. 1997a; Singh et al. 2001; Siddiqui et al. 2004). Some substrates interact directly with sites on the AAA+ conserved core domain. Additional substrate binding sites are located on extra domains that are not shared with other AAA+ proteases. This is a way to ensure that each protease can interact with its own specific substrates.

Variants of ClpX or ClpA missing their specialized domain (the N-domain) are still able to degrade ssrA-tagged proteins, implicating the AAA+ domains in direct substrate recognition (Singh et al. 2001; Dougan et al. 2003; Siddiqui et al. 2004). Accordingly, a point mutation in the pore of ClpX (V154F) within the AAA+ domain is specifically defective in the degradation of C-motif 1 substrates, including the ssrA tag. However, this mutant protein retains the ability to degrade substrates containing each of the other four recognition motifs. These data indicate that there are at least two distinct modes of recognition by ClpX (Siddiqui et al. 2004) (Table 1.2). ΔN-ClpA also retains most of its ability to degrade casein, implicating the core domain in recognition of unfolded proteins (Singh et al. 2001).

Table 1.2. Ability of ClpX variants to interact with different recognition signals.

	C-motif 1	C-motif 2	N-motif 1	N-motif 2	N-motif 3
ClpX	+	+	+	+	+
$ClpX^{\Delta N1-46}$	+	-	+/-	-	-
$ClpX^{V154F}$	-	+	+	+	+

Data based on degradation of Arc reporter proteins fused to representatives from each motif (Siddiqui 2004); (Siddiqui et al. 2004).

\+ = > 50% WT activity
\- = < 50% WT activity
+/- = 50% WT activity

The N-domain of ClpX also appears to play a role in substrate discrimination by ClpXP. ΔN-ClpX is defective in degradation of λO and MuA; ΔN-ClpX is also not able to disassemble MuA transposase complexes in vivo (Table 1.2; Wojtyra et al. 2003). The N-domain of ClpX provides a docking site for all the known adaptor proteins discussed in more detail below (Dougan et al. 2003; Neher et al. 2003b; Bolon et al. 2004). A peptide corresponding to the C-terminal 10 amino acids of the adaptor protein SspB (the XB peptide) binds directly to the N-domain (Bolon et al. 2004). Initial studies indicate that this peptide inhibits the degradation of N-motif 1 and N-motif 2 substrates. It is not yet clear, however, whether the N-domain is mediating direct recognition of these motifs, or if it is involved in their subsequent unfolding and processing steps (S. Siddiqui, personal communication). Further experiments are needed to sort out the different binding sites on ClpX and to find which substrates are competing for the same sites.

No known substrates require the N-domain of ClpA for degradation, however only a small group of substrates have yet been identified. Perhaps the primary function of the N-domain of ClpA is to bind the adaptor protein ClpS, which then mediates the majority of substrate delivery to this protease (Dougan et al. 2002b). The N-domain of Lon appears to play a more global role in substrate selection; N-domain mutants are defective in degradation

of RcsA, SulA, and casein (Roudiak and Shrader 1998). Interestingly, a point mutation in the N-terminal domain of Lon is specifically defective in the degradation of SulA, indicating that there may be distinct substrate binding sites within Lon's N-domain (Ebel et al. 1999). Instead of an N-domain, HslU contains an intermediate (I) domain, located between its two AAA domains. ΔI-HslU is defective in degradation of SulA, but not casein, indicating a substrate discrimination role for this domain (Lee et al. 2003).

These AAA+ proteases must interact with a large assortment of substrates. The variety of signals recognized by each protease can be expanded when the protease carries multiple substrate binding sites. However, this versatility alone does not account for the diversity in substrate recognition. As described below, proteases can also interact with adaptor proteins, some of which themselves can interact with multiple peptide sequences. This modularity rapidly expands the diversity of primary recognition signals that can be recognized by these proteases.

Regulation of proteolysis by cryptic signals.

Although certain proteolytic substrates are constitutively degraded, the hydrolysis of many substrates is temporally coordinated to ensure that they are available when required and are degraded when their functions must be terminated. Intracellular proteases use a number of strategies to guarantee that their substrates are recognized and degraded at the proper time or in response to the correct environmental stimuli (Fig. 1.9). One common approach is to hide the recognition signal until the biologically appropriate time. There are a number of methods used to bury a signal so it is not attainable. For example, recognition sites can be inaccessibly buried in the interior of a folded protein; these signals can be uncovered by unfolding or by cleavage by another protease. Additionally, protein binding partners can mask recognition signals; dissociation of the complex then results in their recognition and destruction.

One of the most basic examples of recognition of latent signals is when the accessibility of a recognition tag changes following denaturation or disassembly of a protein. Native GFP is not a ClpAP substrate, while unfolded GFP is rapidly degraded by this protease (Hoskins et al. 2000b). Quite possibly, unfolding of GFP reveals hydrophobic patches that allow interaction of GFP with ClpA. In certain cases, chaperones such as DnaK and DnaJ have been implicated as adaptor proteins for proteolysis by binding to unfolded proteins and facilitating their binding to ATP-dependent proteases such as Lon and ClpAP, perhaps by maintaining them in a conformation that is susceptible to association with the proteases (Jubete et al. 1996; Huang et al. 2001).

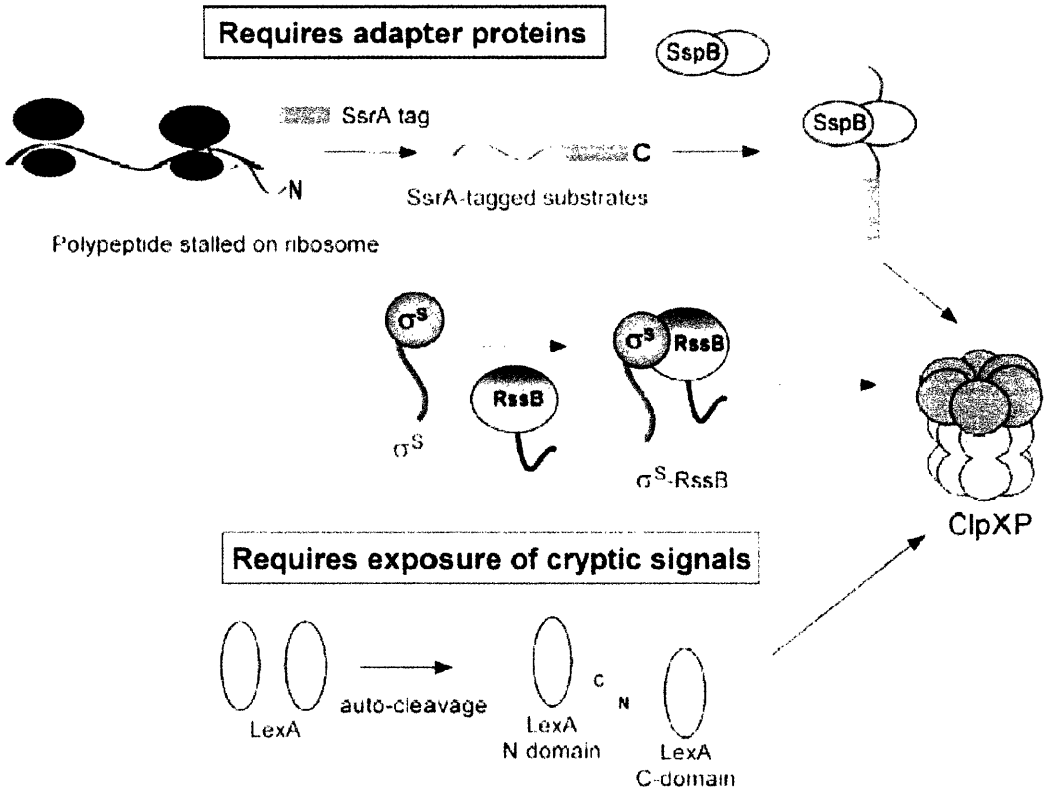


Figure 1.9. Modes of substrate recognition by ClpXP.

LexA is an exquisite example of a protein whose destruction is closely coupled to environmental stress as a result of exposure of a cryptic signal. LexA is the master regulator of the SOS response and undergoes RecA-stimulated self-cleavage following DNA damage (reviewed in Sutton et al. 2000). ClpXP^{trap} captured the fragments generated by autocleavage but not the full-length protein, suggesting that ClpXP recognizes latent signals in LexA that are revealed by this cleavage (Fig. 1.10; see Chapter Three).

Additional experiments determined that ClpXP is able to degrade both the N-terminal and C-terminal fragments of LexA in vitro, but does not degrade the full-length protein (Fig. 1.10b; Neher et al. 2003a). The N-terminal fragment has a newly uncovered C-terminal Val-Ala-Ala sequence which targets it for rapid degradation by ClpXP. Strikingly, this fragment is completely stabilized in *clpX*$^-$ cells, indicating that this protein is solely degraded by ClpXP (Neher et al. 2003b). This specificity could be explained by the fact that ClpXP is the only ATP-dependent protease known to use the free C-terminus of a protein as a primary determinant of recognition. Cleavage of LexA, resulting in positioning of Val-Ala-Ala at the extreme C-terminus, transforms the protein into an attractive ClpXP substrate. Thus, the environmental stimuli, in this case DNA damage, reveals signals in LexA that lead to its degradation by ClpXP, resulting in further activation of the SOS response.

As discussed in Chapter Four, degradation of the N-terminal fragment of RseA is regulated in a similar manner. RseA is the master regulator of the extracytoplasmic stress response. Cleavage of RseA by the protease YaeL reveals a C-motif 1 sequence that targets it for degradation mainly by ClpXP. This recognition of cryptic signals produced following cleavage by another protease is likely a general strategy used by ClpXP to interact with substrates.

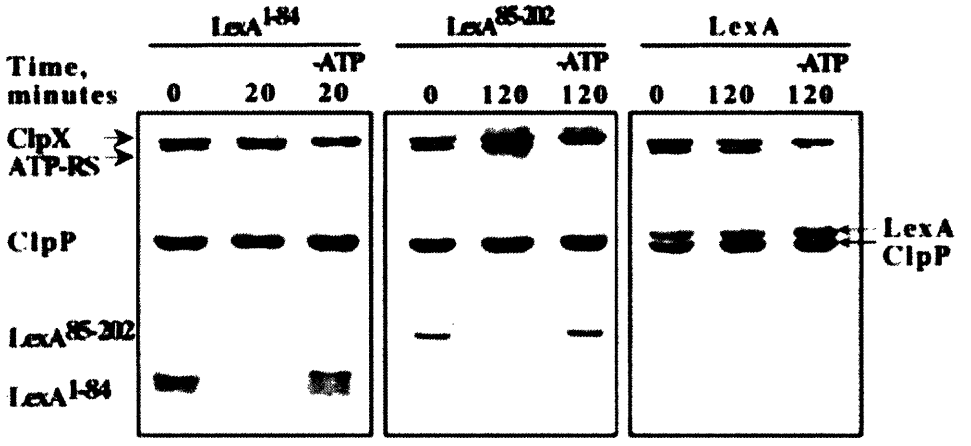


Figure 1.10. LexA is degraded as fragments.

ClpXP-mediated degradation of the N-terminal ($LexA^{1-84}$), C-terminal ($LexA^{85-202}$) and full-length LexA. ClpXP is able to degrade the two auto-cleavage fragments, but not the original full-length protein (figure from Neher et al. 2003a).

Binding partners can also mask signals from recognition by proteases. λO is extremely labile both in vivo and in vitro (Lipinska et al. 1980; Gottesman et al. 1993; Wojtkowiak et al. 1993). To initiate replication of its genome, four dimers of λO bind to the *ori*λ DNA sequence at four repeating sequences, largely inhibiting its degradation by ClpXP (Zylicz et al. 1998). The mechanism of this inhibition remains unknown, however, it is likely that one or more of the λO ClpX-recognition motifs becomes unavailable to ClpX upon assembly of λO on DNA. In addition, many of the ribosomal proteins are rapidly degraded when unassociated and become stabilized upon incorporation into the ribosome. When the L10 ribosomal subunit is over-expressed, it is rapidly degraded; however, it is significantly stabilized upon co-expression of its binding partner L7/L12 (Petersen 1990). Similarly, work described in this thesis shows that in vitro, L10 is rapidly degraded by ClpXP due to a C-terminal signal; however, it is resistant to proteolysis by ClpXP when in complex with L7/L12 (see Appendix

III). Structural studies performed on the L10/L7/L12 complex (the ribosomal "stalk") show that L7/L12 binds to the C-terminal region of L10 (Griaznova and Traut 2000). Together, these data indicate that L7/L12 masks the L10 recognition signal from interaction with ClpXP, allowing the subunit to be stably incorporated into the 50S ribosome. By changing the accessibility of primary ClpXP-degradation tags, the processive destruction of proteins can be elegantly coordinated to the appropriate biological event.

Regulation of proteolysis by adaptor proteins.

Auxiliary specificity factors are a powerful mechanism used to regulate the recognition of proteolytic substrates that have accessible degradation tags (Fig. 1.9). These adaptor proteins can not only modulate substrate choice by proteases, but can also alter the kinetics of substrate binding to enhance interactions at lower substrate concentrations. There are many substrates competing for degradation in the cell; one important role of adaptor proteins is likely to ensure the appropriate degradation of substrates even under low substrate conditions. These factors can also redirect substrate choice by proteases in response to environmental cues, thus playing a critical role in regulating proteolytic flux in the cell. Intracellular proteases can interact with two different types of adaptor proteins: factors that bind to specific substrates and deliver them for degradation, such as SspB, UmuD, and RssB, and those that more generally control the activity of the protease, such as ClpS and MecA.

SspB is a well-characterized ClpX-adaptor protein, identified by its ability to enhance the degradation of ssrA-tagged proteins (Levchenko et al. 2000). Although SspB is not essential for degradation of ssrA-tagged proteins by ClpXP, it greatly increases their rate of degradation at lower substrate concentrations. SspB is a dimer; each SspB monomer contains an N-terminal substrate binding domain, a flexible linker region, and a C-terminal short peptide module (XB) that mediates interactions with ClpX (Fig. 1.11a; Wah et al. 2002; Levchenko et al. 2003).

Mutational studies show that SspB exhibits strong preferences for specific side chains at positions 1-4 and 7 of the ssrA tag (Fig. 1.6). The co-crystal structure of SspB and the ssrA peptide shows that these N-terminal residues of the ssrA tag make specific contacts with the substrate binding domain of SspB, leaving the C-terminal Leu-Ala-Ala amino acids of the tag available for interactions with ClpX (Fig. 1.11b; Levchenko et al. 2003). In this way, ClpX and SspB recognize contiguous portions of the ssrA tag and function in concert to bind ssrA-tagged substrates tightly. The XB modules of SspB bind to the N-terminal domain of ClpX, leashing the substrate to the protease (Bolon et al. 2004). Thus, three weaker protein-protein interactions, those between SspB and ssrA, ClpX and ssrA, and SspB and ClpX, are combined to form an effective delivery complex (Fig. 1.11c, left panel).

Whereas ClpX and SspB recognize contiguous sequences of the ssrA tag, ClpA interacts with determinants that are overlapping those of SspB (Fig. 1.6). As a result, SspB inhibits the recognition and degradation of these substrates by ClpA (see Chapter Two). Thus, SspB is a bifunctional regulator, enhancing recognition of the substrate by one protease, while masking the recognition determinants of a different protease. The ability of SspB to direct ssrA-tagged proteins away from ClpA and towards ClpX for degradation helps explain the observation that although both proteases can degrade these proteins in vitro, ClpXP preferentially degrades them in vivo.

The necessity to re-channel ssrA-tagged proteins towards ClpX was reinforced with the identification of the ClpA-adaptor protein, ClpS (Dougan et al. 2002b). ClpS is a more general adaptor protein that alters the activities of ClpA. ClpS binds to the N-domain of ClpA and inhibits ClpA's recognition of ssrA-tagged proteins, casein, and ClpA itself. In addition, ClpS enhances the ability of ClpA to recognize aggregated proteins (Dougan et al. 2002b; Guo et al. 2002). Thus, both ClpS and SspB inhibit the recognition of ssrA-tagged proteins by ClpA. Why may this redirection of proteolytic activities be important? Perhaps because ClpAP, but not ClpXP, has the general ability to degrade unfolded proteins, and this activity

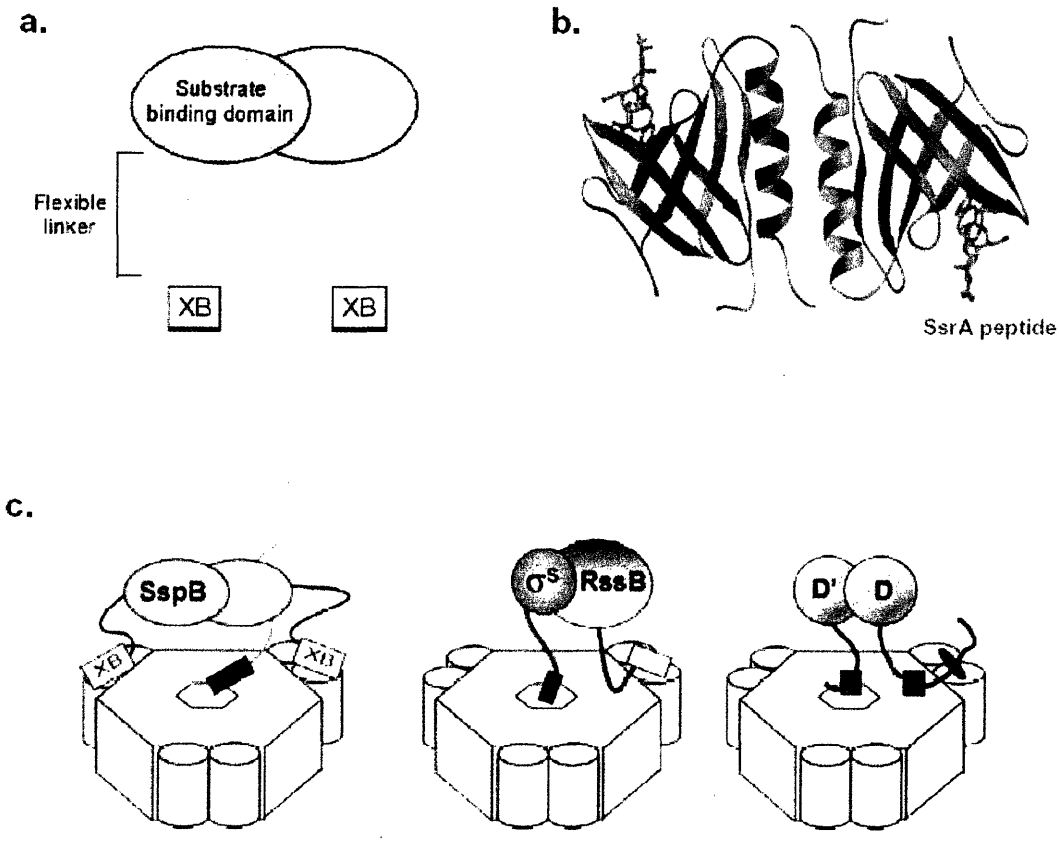


Figure 1.11. ClpXP adaptor proteins.

(a) SspB has an N-terminal substrate binding domain, a flexible linker region, and a C-terminal peptide sequence (XB) that mediates interactions with ClpXP.

(b) Crystal structure of SspB dimer bound to the ssrA peptide (from Levchenko et al. 2003).

(c) Cartoon representations showing various interactions between the adaptor protein, substrate, and ClpX in the following delivery complexes: SspB•ssrA•ClpX (left panel); RssB• σ^S•ClpX (center panel); UmuD•UmuD'•ClpX (right panel) (adapted from Neher et al. 2003b; Bolon et al. 2004).

may be more important than degradation of the ssrA-tagged "cellular trash." These adaptor proteins are re-prioritizing substrate choice by ClpXP and ClpAP so that the most biologically appropriate substrates are degraded. This intricate choreographing of substrate choice could not simply occur by direct recognition of protein signals and requires finer regulation by adaptor proteins.

Are ssrA-tagged proteins the only substrates delivered to ClpXP for degradation by SspB? Capturing substrates in $ClpXP^{trap}$ in strains containing or lacking this adaptor protein indicates that SspB in fact controls the degradation of a number of proteins by ClpXP (see Chapter Four). The N-terminal domain of RseA ($RseA^{1-108}$) was identified as an SspB-dependent ClpXP substrate. SspB enhances the degradation of $RseA^{1-108}$ by ClpXP in vitro in a similar manner as that of the ssrA tag. However, despite the similarities of these interactions, the region of $RseA^{1-108}$ that binds SspB is quite dissimilar from that of the ssrA tag. Thus, although the consensus sequence for interaction of SspB with the ssrA tag has been well defined, the diversity of sequences that can interact with its substrate binding cleft is still not yet understood. The ability of one adaptor protein to deliver multiple targets to a protease efficiently increases the repertoire of proteins that can be recognized and degraded.

Regulator of Sigma S (RssB) is another well-characterized ClpX adaptor protein, required for the degradation of the stationary phase sigma factor, σ^S (Bearson et al. 1996; Muffler et al. 1996; Pratt and Silhavy 1996). σ^S is degraded by ClpXP during exponential-phase growth and is stabilized during stationary phase. RssB has an N-terminal domain homologous to a response regulator and a C-terminal tail that shares similarities with the XB module of SspB. Phosphorylation of RssB regulates its interaction with σ^S; a conserved aspartate in the response regulator domain of RssB becomes phosphorylated in response to unknown signal transduction events during exponential-phase growth (Bouche et al. 1998). Phosphorylated RssB can then bind to an internal sequence of σ^S (Becker et al. 1999). Similar to the SspB•ssrA complex, this RssB•σ^S binary complex can then make two separate

interactions with ClpX: σ^S has an N-terminal degradation signal that binds to ClpX (Flynn et al. 2003; Studemann et al. 2003) and the C-terminal tail of RssB likely binds the N-domain of ClpX (Fig. 1.11c, center panel; S. Siddiqui, personal communication). However, both of these interactions are weak; it is the combination of these recognition elements that results in efficient interaction of σ^S with ClpX only in the presence of phosphorylated RssB, allowing correct regulation of the sigma factor's degradation (Studemann et al. 2003).

UmuD is a ClpXP-adaptor protein that mediates the degradation of UmuD' in a similar manner as the SspB-mediated delivery of ssrA-tagged proteins to ClpXP (Frank et al. 1996). Following DNA damage, UmuD undergoes RecA-stimulated self-cleavage to remove its first 24 amino acids and transform itself into the active form, UmuD' (Shinagawa et al. 1988). UmuD' is a subunit of an error-prone DNA polymerase and its levels are tightly regulated by transcriptional control and degradation (Battista et al. 1990). UmuD' is a substrate for ClpXP, but only when in complex with UmuD (Frank et al. 1996; Neher et al. 2003b). The precursor fragment of UmuD contains a peptide sequence homologous to the XB motif of SspB that tethers the complex to ClpXP to the same site as SspB on the N-terminal domain of ClpX. UmuD' contains one or more weak primary recognition signals that are directly recognized by ClpXP (Fig. 1.11c, right panel; Neher et al. 2003b). Similar to the RssB-σ^S delivery complex, ClpXP regulates the degradation of UmuDD' by combining a tethering interaction with the adaptor protein and a weak primary interaction with the substrate.

SspB, UmuD, and RssB thus share similar mechanisms of substrate delivery to ClpXP. Multiple weak protein-peptide interactions between the adaptor protein, substrate, and protease are combined to enhance the specificity and affinity of degradation (Fig. 1.11). The fact that these multiple adaptor proteins appear to bind to the same tethering site on the N-terminal domain of ClpX indicates that increasing the levels of a certain adaptor protein in response to environmental stimuli could re-prioritize the degradation of certain substrates under these conditions. In addition, as seen above, certain substrates may be also binding to

this site on ClpX. Competition between substrates and adaptor proteins for the same sites on ClpXP would add an additional layer of regulation of substrate choice by this protease.

Whereas the known ClpXP adaptor proteins – SspB, RssB, and UmuD – appear specialized for certain substrates, adaptor proteins for other prokaryotic proteases have a more global effect on the activities of their cognate enzymes. For example, MecA is an adaptor protein that associates with the *B. Subtilis* Hsp100/Clp protein, ClpC (Turgay et al. 1997). MecA is a dimeric protein with an organization similar to SspB; an N-terminal substrate binding domain and a C-terminus that binds ClpC (Persuh et al. 1999). However, MecA targets a large variety of substrates to ClpCP for degradation including ComK, ComS, casein, and aggregated proteins (Turgay et al. 1997; Schlothauer et al. 2003). In fact, studies indicate that MecA is essential for activation of ClpC and that ClpC cannot recognize substrates on its own (Schlothauer et al. 2003). Coupling activation of ClpCP to substrate delivery is an efficient strategy to regulate degradation.

Adaptor proteins thus play a critical role in controlling proteolytic flux in the cell. The presence or absence of an adaptor protein can have great affects on the repertoire of substrates degraded by a protease. For instance, the set of substrates captured by $ClpXP^{trap}$ changes depending of the presence of the adaptors SspB and RssB (see Chapter Four and Appendix III). The ability of each protease to recognize multiple adaptor proteins and each adaptor protein to recognize multiple substrates expands the number of proteins that can be recognized by a single protease, and provides the opportunity for both regulation and competition.

Spatial regulation of degradation by ClpXP.

The strategies of controlling degradation discussed thus far have been temporal regulation. However, another effective method is to spatially restrict the degradation of a substrate to a specific location in the cell. For example, Lon protease is known to bind to

DNA, and thus Lon may be targeted to degrade certain substrates specifically when both are associated with DNA (Fu et al. 1997; Fu and Markovitz 1998). The membrane protease, FtsH, is advantageously located to degrade inner membrane proteins (Tomoyasu et al. 1995). The ClpXP adaptor protein, SspB, was initially identified as a ribosome-associated protein. This may assist in enhancing degradation of ssrA-tagged proteins by localizing the adaptor protein to the same address as the substrate (Levchenko et al. 2000).

The intrinsic asymmetric nature of *C. crescentus* cell division makes this bacterium a model organism to study the role of polar localization in cell cycle progression. Each cell division is asymmetric, giving rise to two morphologically distinct progeny; a non-motile stalked cell and a motile swarmer cell (Fig. 1.12a). CtrA is an essential transcriptional regulator that controls cell cycle progression in this bacterium (Quon et al. 1996). CtrA's activities are controlled by transcription, proteolysis by ClpXP, and phosphorylation, so that its active form is eliminated at the swarmer-to-stalked cell transition immediately before the cell begins to replicate its DNA and is most abundant pre-cell division (Fig. 1.12b; Domian et al. 1997; Domian et al. 1999). CtrA accumulates at the cell pole just before its proteolysis; this proteolysis depends on correct localization (Ryan et al. 2002). However, ClpXP is active throughout the cell cycle, indicating that a mechanism is required to regulate CtrA's correct destruction (Jenal and Fuchs 1998).

CtrA has a bipartite recognition signal that is sufficient for its cellular localization and degradation by ClpXP (Ryan et al. 2002). The C-terminal Ala-Ala residues are necessary for its degradation but not localization (Domian et al. 1997) whereas a signal in the N-terminal 56 amino acids is required for both its degradation and localization (Ryan et al. 2002). Because CtrA is only degraded when it is correctly localized, one possibility is that there is an unidentified adaptor protein that recruits CtrA to the pole and promotes its interaction with ClpXP. This localization of CtrA's proteolysis ensures that it is only degraded in the stalked

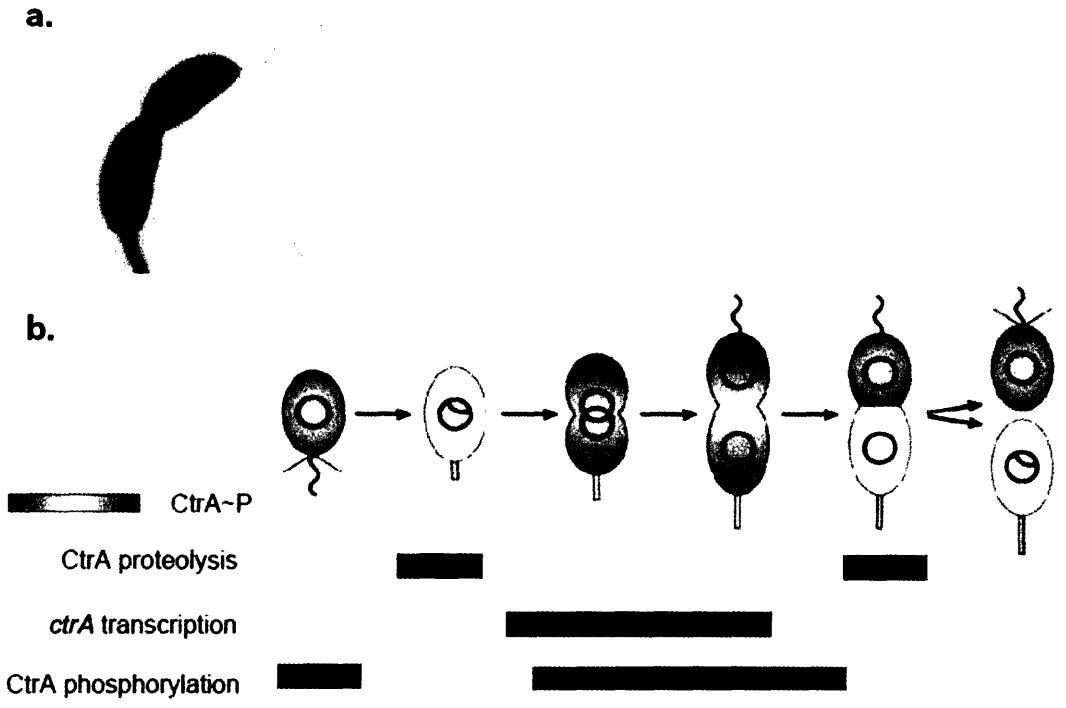


Figure 1.12. Regulation of the master cell cycle regulator CtrA in *C. crescentus.*

(a) Electron micrograph of a *C. crescentus* predivisional cell (Skerker and Laub 2004).

(b) CtrA's activities are controlled by proteolysis, transcription and phosphorylation, so that the active form, CtrA~P, is eliminated at the swarmer-to-stalked cell transition before the cell begins to replicate its DNA and is most abundant pre-cell division (figure adapted from Skerker and Laub 2004).

half of the pre-divisional cell, whereas ClpXP is active throughout the cell (Ryan et al. 2002; Ryan et al. 2004). This is an exquisite example of a protein whose correct degradation is mediated by a specific sub-cellular address.

Combinatorial recognition of multiple signals.

Intracellular ATP-dependent proteases thus have an arsenal of recognition strategies they use in a combinatorial fashion to regulate protein degradation. Substrate recognition involves multiple weak protein-peptide interactions. As seen in the first section of this introduction, most, if not all substrates have primary recognition motifs encoded in their sequences. In some cases, substrates have more than one signal; in other cases, an interaction between an adaptor protein and the protease is combined with the protease's direct interaction with the primary degradation signal. This coupling of multiple weak interactions provides a number of combinatorial regulatory advantages. Recognition of one weak interaction is not enough to fully engage the substrate with the protease; thus, for instance, recognition of one signal could depend on the availability of the other signal. The second signal could be masked in a protein-protein interaction, could be available only upon a protein processing event, or could be provided by an adaptor protein whose availability is regulated. This combination of multiple signals ensures the appropriate biological proteolytic response and permits regulation of the proteome by degradation.

The critical role proteolysis plays in regulating the proteome is extremely apparent in the post-translational control of the levels and activity of sigma factors. *E. coli* RNA polymerase is composed of the core enzyme ($\alpha_2\beta\beta'$) and one of seven different sigma subunits; each sigma factor binds to a specific set of promoter regions allowing expression of a defined set of genes (Helmann and Chamberlin 1988; Ishihama 1988; Gross et al. 1998). Degradation is known to play an important role in the activation of three of these sigma factors, σ^S, σ^H, and σ^E. It is essential that these sigma factors are degraded only under the

appropriate biological conditions; the proteases that regulate the availability of these sigma factors combine the recognition of multiple degradation signals with a variety of regulatory strategies to ensure their proper degradation.

σ^S (σ^{38}): σ^S is responsible for transcription of more than 50 genes whose products are involved in stationary phase growth (reviewed in Hengge-Aronis 1996a; Hengge-Aronis 1996b; Ishihama 2000). The increased abundance of σ^S in stationary phase is mainly due to its enhanced resistance to degradation by ClpXP under these conditions (for review see Hengge-Aronis 1996a). σ^S has an N-terminal ClpX-recognition signal, however, this signal is not sufficient to engage degradation. To correctly destroy this protein and thus regulate its availability, ClpXP combines the recognition of this signal in σ^S with interaction with the adaptor protein RssB whose availability changes in response to environmental stimuli.

σ^H (σ^{32}): σ^{32} is the heat shock transcription factor. σ^{32} is stabilized against degradation by FtsH following heat shock, allowing it to respond to this stress (Herman et al. 1995; Tomoyasu et al. 1995). The alteration of molecular binding partners regulates the appropriate degradation of σ^{32} by FtsH. The mechanism of degradation of σ^{32} is not completely understood, however, DnaK and DnaJ chaperones appear to play a role, since σ^{32} is stabilized in *dnaK* and *dnaJ* cells. The hypothesis is that during normal growth conditions, DnaK binds to σ^{32} and enhances its availability to FtsH (Tilly et al. 1989; Straus et al. 1990; Tomoyasu et al. 1998). The N-terminal region of σ^{32} appears important for its recognition by FtsH (Nagai et al. 1994; Tomoyasu et al. 2001), however, it is also possible that FtsH interacts with σ^{32} under these conditions due to its folded state instead of sequence specificity (Herman et al. 2003). During heat shock, DnaK is needed to bind and prevent the aggregation of misfolded proteins, releasing σ^{32} so it is available to bind the core RNA polymerase (reviewed in Yura 2000), preventing its degradation by FtsH (Tomoyasu et al. 1998). Although the

mechanism of recognition of σ^{32} by FtsH is not completely understood, it is clear that a number of strategies are used to ensure its swift availability during heat shock. σ^{32}'s interaction with a protein binding partner, DnaK, increases its recognition by FtsH. In addition, interaction with another binding partner, the core RNAP, masks this degradation signal.

σ^E (σ^{24}): σ^E is an essential sigma factor that controls the expression of a set of genes that cope with periplasmic stresses such as misfolded proteins (Erickson and Gross 1989; Raina et al. 1995; Rouviere et al. 1995). The cytoplasmic abundance of σ^E is highly regulated mainly through proteolysis of its anti-σ factor, RseA. Periplasmic stress signals are relayed across the inner membrane by a proteolytic cascade mediated through RseA (Alba et al. 2001; Alba et al. 2002; Kanehara et al. 2002), releasing N-RseA into the cytoplasm (Alba et al. 2001; Alba et al. 2002; Kanehara et al. 2002). This cleavage of RseA reveals a previously hidden ClpXP recognition signal. SspB enhances degradation of N-RseA by ClpXP, likely giving N-RseA a competitive edge over other substrates. This is an elegant case of the use of multiple recognition strategies – cryptic signals and adaptor proteins – to couple environmental change to the degradation of the appropriate substrate.

These examples illustrate the power of proteolysis as a method to couple changes in the environment to changes in gene expression. In each case, the accessibility of a primary degradation signal in a substrate to its respective protease is regulated to ensure correct proteolysis. Using a variety of regulatory mechanisms, proteases can rapidly adjust the concentration of transcription factors in response to the correct stimuli.

Recognition of peptide signals as a common means of regulation.

As a picture of signal recognition by bacterial intracellular proteases becomes more apparent, one emerging theme is their ability to recognize highly localized information in short

peptide sequences. A few peptide sequences studied are amazingly rich in recognition sequences, with short peptides mediating multiple protein-protein interactions. For example, the 11 amino acid ssrA tag encodes information recognized by at least five proteins – ClpX, ClpA, SspB, FtsH, and Tsp (Keiler and Sauer 1996; Gottesman et al. 1998; Herman et al. 1998). The residues that mediate interactions with ClpX, ClpA, and SspB are distinct, overlapping sets of sequences; differential recognition of the ssrA tag by these three proteins, depending on their availability, may reflect their ability to redirect substrate degradation during various conditions (see Chapter Three). The C-terminal region of N-RseA also appears to be mediating a number of protein-protein interactions. Following cleavage by YaeL, the C-terminal residues of N-RseA are recognized by SspB, ClpX and likely a number of other proteases that all work together to destroy the protein and release σ^E (see Chapter Four).

There are many other biological examples of peptide sequences embedded within proteins that mediate multiple protein-protein interactions. These peptide signals can modulate numerous aspects of the attached protein's cellular fate such as localization and binding partners. For example, peptide signals often control a protein's final cellular address. In addition, control in signal transduction cascades are often regulated by recognition of peptide motifs by a variety of peptide binding domains.

One compelling example of a peptide richly encoded in protein interaction information is that of the signal peptide. Many proteins destined for the periplasm or outer membrane harbor cleavable N-terminal signal peptides that target these proteins to the transport apparatus by the SecA pathway (reviewed in Muller et al. 2001; also see Krieg et al. 1986; Lehnherr and Yarmolinsky 1995). Signal peptides, such as that of pro-OmpA, are typically 18-26 amino acids in length and possess a common structure containing a positively charged N-terminal region, a central hydrophobic region, and a C-terminal cleavage site for the signal peptidase (Fig. 1.13). This pro-peptide makes many different protein contacts on its journey from the ribosome to the outer membrane. Trigger Factor (TF) interacts with the signal

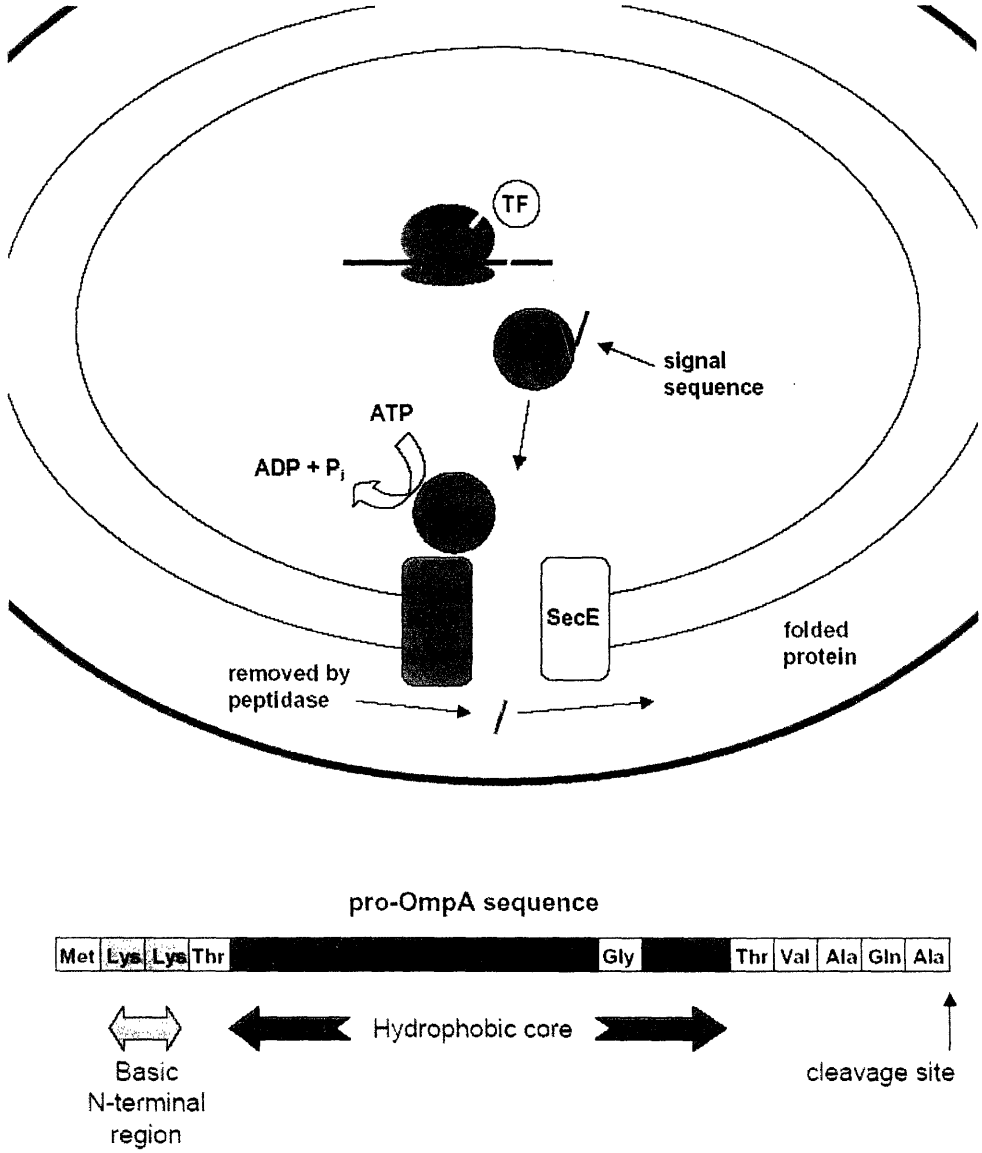


Figure 1.13. Model of the localization and processing of a pro-protein.

The signal sequence of a pro-protein makes multiple protein contacts on its pathway to becoming correctly processed and localized, including with Trigger Factor (TF), SecA, and the signal peptidase.

sequence as it emerges from the ribosome (Eisner et al. 2003). TF recognizes a motif of eight residues favoring basic and hydrophobic residues (Patzelt et al. 2001; Deuerling et al. 2003). This TF-binding motif overlaps that of DnaK, which consists of a hydrophobic core of five residues flanked by basic residues (Rudiger et al. 1997; Deuerling et al. 2003). DnaK may function to rescue misfolded proteins downstream of TF (Deuerling et al. 2003). Following these contacts, SecA then binds to the leader peptide sequence targeting pro-OmpA to the translocation pore where it is secreted and recognized and cleaved by the signal peptidase (Fig. 1.13; for review see Driessen et al. 1998).

Intriguingly, pro-OmpA was also captured by $ClpP^{trap}$, indicating that it is a ClpXP substrate (see Chapter Three; A. Abdelhakim, unpublished data). Pro-OmpA, along with other secreted proteins, has an N-terminal ClpX-recognition signal that overlaps with the signal peptide. One attractive model is that pre-proteins that are not correctly excreted become substrates for degradation due to recognition of their mislocalized secretion tag. Thus, the signal peptide of pro-OmpA mediates many overlapping protein-peptide interactions whose correct recognition regulate the proper localization of this protein.

The specificity in signal transduction is mediated primarily through protein-protein interaction domains. Combinatorial recognition of primary-sequence motifs by an array of modular domains define the structure of complex signaling networks that control virtually every aspect of cellular function. SH2 domains are the prototype of modular protein domains found in signaling molecules. SH2 domains recognize short motifs consisting of a phosphorylated tyrosine (pTyr) residue along with three to six C-terminal residues; discrimination in peptide binding is largely determined by the sequence surrounding the pTyr (Fig. 1.14a; Songyang et al. 1993; Songyang et al. 1994). Other peptide binding modules involved in cell signaling pathways include the PDZ domain. These domains have been implicated in a variety of protein associations including coupling of receptors to enzymes (for

review see Fanning and Anderson 1996). PDZ domains bind short peptide sequences with a C-terminal hydrophobic residue and a free carboxylate (Songyang et al. 1997).

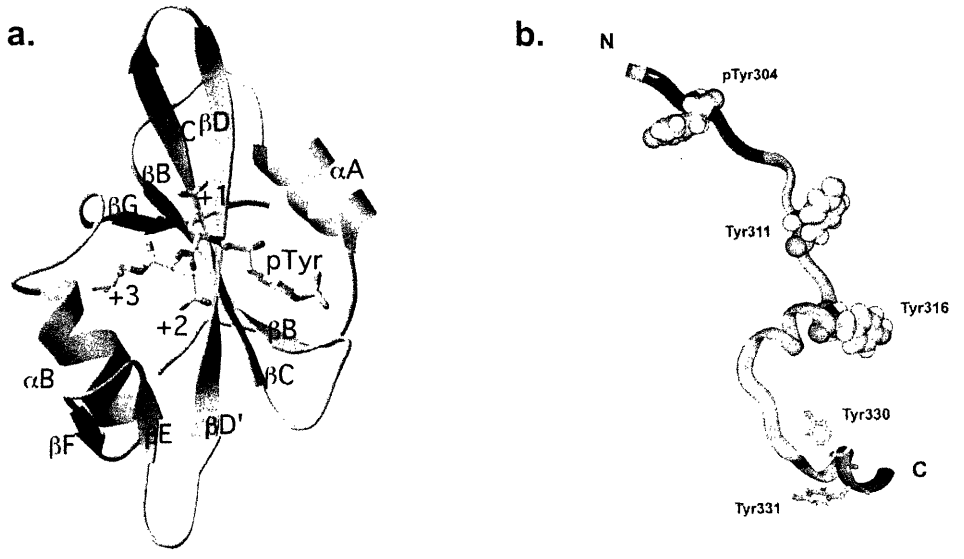


Figure 1.14. SH2 domains recognize short peptide motifs containing a pTyr residue.

(a) Schematic ribbon diagram of an SH2 domain. The structure shown is the Src SH2 domain complexed with hmT (hamster middle T antigen) phosphopeptide. The phosphotyprosine (pTyr), glutamate (+1), glutamate (+2), and isoleucine (+3) of the hmT peptide are shown. The SH2 domain structure consists of a large β-sheet flanked by two α-helices. The pTyr in the phosphopeptide inserts into a positively-charged pocket in the SH2 domain located on the N-terminal side of the central β-sheet (Kuriyan and Cowburn 1997).

(b) Schematic diagram of the ephrinB2 (301-333)-pY304 peptide. Residues for the Grb4 SH2 domain binding, i.e. PHpY304EKV, are colored in blue, while the tail residues IY330YKV for PDZ domain binding are colored in red (Su et al. 2004).

SH2 binding specificity was elegantly assayed by probing a degenerate peptide library containing a pTyr residue against a number of SH2 domains. The binding selectivity of the SH2 domains in the in vitro selection experiment correlated well with known in vivo preferences (Songyang et al. 1993; Songyang et al. 1994). Importantly, this screen identified amino acids that are both favored and disfavored; both of these forces likely contribute to binding specificity in the cell. These defined binding motifs could then be applied to bioinformatic techniques to search for new interacting partners of specific SH2 domains (Yaffe et al. 2001). This is a powerful technique that can be applied to many different peptide binding domains, including that of the intracellular proteases.

Due to the highly concise nature of these recognition motifs, peptides can be embedded with motifs that mediate the association with multiple peptide binding domains. For example, the B class of ephrins are membrane-bound ligands that transduce a "forward" signal to cells expressing Eph receptors, and "reverse" signals to the cell expressing the ephrins (for review see Schmucker and Zipursky 2001; Cowan and Henkemeyer 2002). These ephrin/Eph signaling pathways mediate many cell-cell communications, for example those required for axon guidance (for review see Kullander and Klein 2002). The 33 amino acid C-terminal tail of ephrin B carries binding motifs for at least two independent docking proteins (Fig. 1.14b). Phosphorylation of a tyrosine within this tail confers binding of ephrin B2 to the SH2 domain of Grb4. In addition the PDZ domain of PDZ-RGS3 can bind this peptide independently of phosphorylation (Cowan and Henkemeyer 2001; Lu et al. 2001). In fact, GST pull-down experiments indicate that ephrin B2 can simultaneously mediate both of these interactions (Su et al. 2004). Thus, this short C-terminal tail of ephrin is rich in signaling information, encoding sequences that control downstream networks, mediating specific signaling events.

These examples demonstrate the ability of short peptide sequences to regulate many facets of a protein's activities. Similar to the motifs directly recognized by bacterial proteases,

many different mechanisms can be used to regulate the availability of these peptide sequences to binding partners. The multitude of techniques that have been developed to study the specificity of these peptide sequences for their binding partners will likely provide a wealth of information on specificity in recognition by ClpXP.

Recognition by the proteasome in higher organisms.

The eukaryotic 26S proteasome uses a unique mechanism to select its substrates; proteins targeted for degradation by the proteasome are first covalently modified by poly-ubiquitin molecules. Regulated proteolysis by the ubiquitin system plays an essential role in many cellular processes, including cell cycle, stress responses, and development, just to name a few. The list of proteins targeted by ubiquitin is rapidly growing (for review see Glickman and Ciechanover 2002). It is clear that although the mechanism of direct substrate recognition between the prokaryotic and eukaryotic systems are quite different, they employ many of the same mechanisms to regulate this recognition.

Degradation of substrates by the proteasome occurs first by covalently tagging a protein with ubiquitin, and then by recognition and degradation of the ubiquitinated substrate by the proteasome. Conjugation of a protein by ubiquitin (Ub) occurs in a 3-step cascade: Ub is first activated by formation of a thioeseter bond with the Ub-activating enzyme (E1). Ub is then transferred to one of a large family of Ub-conjugating enzymes (E2). E3 Ub-protein ligases bind to the substrate and the activated E2 and mediate the transfer of the Ub molecule from E2 to the substrate. Thus, E3's play the key role in substrate selection because they are responsible for choosing specific proteins for ubiquitination (for review see Hochstrasser 1996; Hershko and Ciechanover 1998).

Similar to the bacterial intracellular proteases, many regulatory strategies are employed by the Ub-conjugating system to ensure the correct degradation of substrates. The ability of this system to target such a diverse array of substrates arises mainly from the

modularity of the system. There are 11 different E2 enzymes (in yeast), each of which can interact with several of many E3s. In turn, specific E3s can often interact with several different substrates by the same or different recognition motifs. Additional complexity is added to this hierarchal structure when, for example, one substrate interacts with more than one E3, or one E3 can interact two different E2s, etc. The large number of complexes that can be formed greatly expands the repertoire of proteins that can be specifically degraded by the proteasome (for review see Glickman and Ciechanover 2002). Delving into the complexity of regulation by the Ub system is beyond the scope of this introduction, however, a few examples will give a flavor of the diverse recognition strategies used.

In eukaryotes, N-end rule substrates – proteins with destabilizing N-terminal amino acids – bind directly to an E3 ligase via their N-terminal residue and are subsequently degraded by the proteasome (Fig. 1.8b; Varshavsky 1992). This E3, Ubr1, has three substrate recognition sites, one for substrates with basic N-terminal residues, one for hydrophobic N-terminal residues, and one that binds non-N-end rule substrates (Reiss et al. 1988; Kwon et al. 1998). Thus, this one enzyme can target substrates with three different recognition motifs for degradation; this is a clear example of how the modularity of the Ub-conjugating system significantly expands the substrate repertoire of the proteasome.

It is easy to see how recognition of proteins based on the nature of their N-terminal amino acid would lend well to targeting the degradation of cleaved proteins. As we've seen with ClpXP, revealing hidden recognition signals by a site-specific cleavage is an elegant mechanism to temporally control proteolysis. One example of the N-end rule substrates demonstrates the utility of this mechanism in regulating proteolytic substrate recognition. A cohesin complex, containing a protein called SCC1, holds chromatids together. During anaphase, the site-specific protease, Esp1, is activated and cleaves SCC1 releasing three fragments (Uhlmann et al. 2000). One of these fragments contains a newly revealed N-terminal arginine residue that targets it for degradation in a Ubr1-dependent manner (Rao et

al. 2001). Cleavage and degradation of SCC1 allows for separation of the sister chromatids. Thus, selective degradation of the SCC1 fragment occurs in a temporally regulated fashion.

A couple of different mechanisms are used to regulate the proteolysis of the α2 mating factor in yeast. Haploid yeast cells express one of two transcription factors, a1 or α2, that dictates the mating type. Both of these factors are degraded rapidly in a Ub-dependent manner. When two haploid cells of different types mate, these proteins are stabilized, and the resulting diploid cell expresses both a1 and α2. α2 contains two degradation signals, Deg1 and Deg2, each recognized by a different set of E2/E3 pairs. Additional complexity arises from the fact that two E2's work with one E3 to recognize Deg1, while two different E2's associate with a different E3 to recognize Deg2 (Chen et al. 1993). This is an amazing example of the combinatorial use of Ub-conjugating enzymes with overlapping substrate specificities to ensure the correct and efficient degradation of a substrate. Stabilization of α2 in the diploid cell results from masking of these degradation signals. Deg1 consists of hydrophobic residues on one face of an α-helix; this determinant overlaps with the residues important for heterodimerization with a1. Thus, complex formation between α2 and a1 stabilizes the protein against degradation by hiding Deg1 from the degradation machinery (Johnson et al. 1998).

Although the bacterial intracellular proteases and the proteasome have quite disparate methods to target their substrates for degradation, it is clear that common modes of regulation of these processes have evolved. Both systems combine a range of recognition strategies, such as direct substrate recognition and the masking of signals, to broaden their substrate repertoire while still using discretion in selecting substrates for degradation.

Summary

The degradation of cellular proteins by AAA+ proteases is a highly complex and tightly regulated process. ClpXP interacts with a broad range of substrates by directly recognizing different classes of primary sequence motifs. In addition to this direct recognition, ClpXP uses a variety of regulatory strategies to correctly target its substrates for degradation. Adaptor proteins enhance the affinity of certain substrates for the protease. Hidden recognition signals can be exposed in response to the correct environmental stimulus. It is likely that as new ClpXP substrates are discovered, novel mechanisms this protease uses to correctly select its substrates will be revealed.

This dissertation focuses on these various strategies used by ClpXP to correctly select its substrates for recognition. Chapter Two dissects the overlapping sequences in the ssrA tag recognized by the two proteases ClpXP and ClpAP, and the ClpXP adaptor protein, SspB, by mutational analysis. These experiments led to an enhanced understanding of the binding motifs recognized by each of these proteins and provided insight into how they can function together to correctly degrade ssrA-tagged proteins. In Chapter Three, we identified many new ClpXP substrates using an in vivo trapping method. This identification allowed us to define five classes of ClpXP-recognition motifs based on sequence similarities, greatly increasing our understanding of direct substrate recognition by ClpX. Finally, in Chapter Four, we used a similar in vivo trapping technique to identify additional substrates that rely on SspB to be targeted to ClpXP. The N-terminal fragment of RseA is one of these substrates; SspB enhances the degradation of N-RseA both in vitro and in vivo. Thus, a combination of recognition strategies – a hidden signal and adaptor protein – is used to regulate the correct degradation of this substrate.

CHAPTER TWO:

Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis

---

This chapter was previously published as Flynn, J.M., I. Levchenko, M. Seidel, S.H. Wickner, R.T. Sauer, and T.A. Baker. *Proc Natl Acad Sci U S A* **98**: 10584-9 (2001). I. Levchenko performed the filter binding experiment in Figure 2.2 and purified SspB, M. Seidel cloned most of the GFP-ssrA mutants, and S. Wickner provided the purified ClpA. R.T. Sauer and T.A. Baker assisted in preparation of the manuscript.

Abstract

The ssrA tag, an 11 amino-acid peptide added to the C-terminus of proteins stalled during translation, targets proteins for degradation by ClpXP and ClpAP. Mutational analysis of the ssrA tag reveals independent, but overlapping determinants for its interactions with ClpX, ClpA, and SspB, a specificity-enhancing factor for ClpX. ClpX interacts with residues 9-11 at the C-terminus of the tag, whereas ClpA recognizes positions 8-10 in addition to residues 1-2 at the N-terminus. SspB interacts with residues 1-4 and 7, N-terminal to the ClpX binding determinants, but overlapping the ClpA determinants. As a result, SspB and ClpX work together to recognize ssrA-tagged substrates efficiently, whereas SspB inhibits recognition of these substrates by ClpA. Thus, dissection of the recognition signals within the ssrA tag provides insight into how multiple proteins function in concert to modulate proteolysis.

Introduction

The proteolytic machinery of cells must select the correct protein substrates at the right time and place. Two general mechanisms, degradation tags and regulatory proteins that modulate recognition, help ensure intracellular proteolytic specificity. Degradation signals, which can be present in the protein sequence or added by covalent modification, target substrates to specific proteases. In eukaryotes, for example, proteins can be targeted to the 26S proteosome by post-translational addition of polyubiquitin (Hochstrasser 1996; Hershko and Ciechanover 1998; Voges et al. 1999). In bacteria, proteins bearing the ssrA degradation tag, an 11-residue peptide, are recognized and degraded by several different proteases, including ClpXP and ClpAP (Gottesman et al. 1998). The ssrA tag is added cotranslationally to the C-terminus of polypeptides whose biosynthesis has stalled (Tu et al. 1995; Keiler et al. 1996; Roche and Sauer 1999). The specificity of proteolysis can be further regulated by protein factors that modulate recognition of degradation signals by the protease. In *E. coli*, for example, the SspB protein binds specifically to ssrA-tagged substrates and enhances binding of the tagged protein to ClpX (Levchenko et al. 2000).

ClpXP and ClpAP are protein machines that promote ATP-dependent degradation. Each of these complexes contains a hexameric Clp/HSP100-family ATPase, ClpA or ClpX, that mediates substrate recognition and catalyzes energy-dependent protein unfolding (Gottesman et al. 1993; Wojtkowiak et al. 1993; Wickner et al. 1994; Levchenko et al. 1995; Wawrzynow et al. 1995; Weber-Ban et al. 1999; Hoskins et al. 2000b; Kim et al. 2000). Both Clp ATPases can form a stacked protease complex with ClpP, a double-ring serine peptidase whose active sites face an internal chamber (Wang et al. 1997; Gottesman et al. 1998; Grimaud et al. 1998; Ortega et al. 2000). The entrance to the inner proteolytic compartment of ClpP is small (Wang et al. 1997) and, prior to degradation, substrates must be unfolded by ClpX or ClpA and translocated into ClpP. Although similar in function and in their ability to

recognize ssrA-tagged proteins, ClpX and ClpA have generally distinct substrate preferences. For example, ClpXP degrades the stationary-phase sigma factor (Schweder et al. 1996) and Mu transposase (Levchenko et al. 1995) which are not substrates for ClpAP, whereas ClpAP, but not ClpXP, degrades HemA (Wang et al. 1999) and MazE (Engelberg-Kulka and Glaser 1999). Moreover, ClpAP, but not ClpXP, degrades denatured proteins in the absence of a degradation tag (Katayama et al. 1988; Hoskins et al. 2000a). Although a few specific recognition sequences for ClpX and ClpA have been identified, general sequence rules governing substrate recognition by either protein have yet to emerge.

In this paper, we determine the sequence information within the ssrA degradation tag that is required for efficient recognition by ClpX, ClpA, and SspB. We find that the ssrA tag is rich in signaling information. ClpX and SspB recognize contiguous portions of the ssrA tag, and function in concert to bind ssrA-tagged substrates tightly, allowing more efficient degradation of these substrates by ClpXP. In contrast, SspB interacts with sequence determinants that partially overlap those of ClpA, resulting in inhibition of ClpAP-mediated degradation. These results establish that SspB can act as a bifunctional regulator of substrate recognition and that the ssrA tag contains intricate, overlapping recognition signals that allow modulation of proteolysis.

Results

Mutant derivatives of the ssrA tag.

To identify the residues within the 11 amino-acid ssrA tag that are important for recognition by ClpX and ClpA, we constructed a set of mutant tags fused to the C-terminus of green fluorescent protein (GFP-ssrA). Each non-alanine residue in the tag sequence was mutated to alanine, and each alanine was changed to aspartic acid (Fig. 2.1A). Because GFP-ssrA (L9A) (numbering relative to the N-terminus of the tag) was a relatively conservative mutation, we also constructed the GFP-ssrA (L9D) mutant. In total, twelve single GFP-ssrA mutants were constructed and purified.

To assay recognition by ClpX and ClpA, we measured degradation of the GFP-ssrA variants by ClpXP and ClpAP *in vitro*. The initial rate of degradation of each mutant was determined by measuring the loss of GFP-ssrA fluorescence. To determine K_m values, degradation rates were determined at a series of substrate concentrations. Consistent with previous reports, K_m for ClpXP degradation of GFP-ssrA was 1.5 ± 0.3 μM (Kim et al. 2000; Levchenko et al. 2000) (see Fig. 2.1C). K_m for ClpAP degradation of GFP-ssrA (1.5 ± 0.4 μM) was found to be similar (see Fig. 2.1E).

SsrA-tag•ClpX recognition.

Of the twelve GFP-ssrA mutants tested, only those with substitutions at tag positions 9, 10 and 11 caused greater than 2-fold increases in K_m for ClpXP degradation relative to the wild-type value (Fig. 2.1B). Ala^{10} and Ala^{11} were found to be critical determinants for recognition by ClpX. GFP-ssrA with either the A10D or A11D substitution had a K_m for ClpXP degradation that was increased by at least a factor of 100 (no degradation was observed at substrate concentrations of 100 μM). Mutation of Leu^9 to either alanine or aspartic acid also weakened productive interaction of the substrate with ClpX, increasing the K_m about four-fold

(L9A K_m = 6.2 ± 0.6 μM; L9D K_m = 6.9 ± 1.1 μM). In contrast, residues 1-8 of the ssrA tag did not play major roles in ClpX recognition, as judged by K_m values similar to wild-type. Furthermore, V_{max} values for the mutants with detectable degradation rates were similar to the wild-type value of 1.2 ± 0.1 min^{-1} $ClpX_6^{-1}$ (data not shown, but see legend to Fig. 2.1). Our finding that Ala^{10} and Ala^{11} play the largest role in ClpX recognition is consistent with previous studies showing that replacing both residues with aspartic acids greatly reduces degradation by ClpXP of a tagged version of the N-terminal domain of λ repressor (Gottesman et al. 1998).

To determine whether the Leu^9-Ala^{10}-Ala^{11} sequence motif was sufficient to mark a protein as a substrate for ClpX, we constructed two additional variants. In one protein, residues 1-8 of the tag were mutated to the same amino acids shown in Fig. 2.1A to generate GFP-D_2A_5DLAA. In the other, residues 1-8 of the tag were changed to glycines resulting in GFP-G_8LAA. The GFP-D_2A_5DLAA protein was a substrate for ClpXP degradation (Fig. 2.1C), although with an increased K_m value (10.1 ± 1.4 μM). This change in K_m probably results from the cumulative minor effects of the eight single mutations. The glycine-rich GFP-ssrA variant was resistant to degradation by ClpXP at concentrations of 50 μM and below (data not shown). We conclude that a C-terminal Leu-Ala-Ala tripeptide is sufficient to allow ClpX recognition and ClpXP-dependent degradation in some but not all sequence contexts. Because of its flexibility, the glycine-rich linker may not allow the terminal Leu-Ala-Ala residues to adopt a conformation appropriate for ClpXP recognition.

SsrA-tag•ClpA recognition.

Degradation of the GFP-ssrA mutants by ClpAP (Fig. 2.1D) revealed that ClpA relies on a different set of residues than ClpX to recognize the ssrA tag. The mutations that caused the largest increases in K_m for ClpAP degradation (wild-type value 1.5 μM) were: A1D (14.3 ± 1.5 μM), A2D (6.4 ± 1.5 μM), A8D (10.1 ± 1.7 μM), L9D (17.1 ± 1.2 μM), and A10D (4.5 ± 0.4

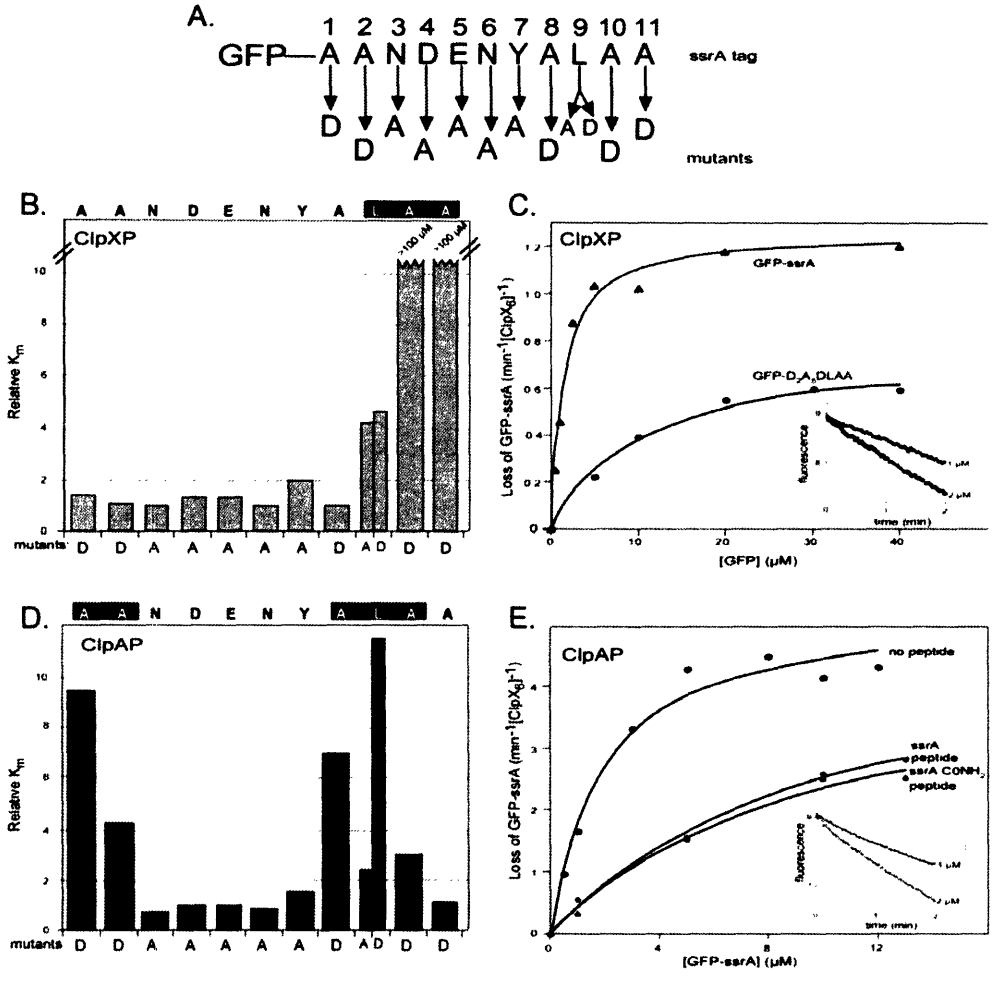

A.
1 2 3 4 5 6 7 8 9 10 11
GFP—A A N D E N Y A L A A
ssrA tag
D A A A A A D D D
D A A D D
mutants
B.
A A N D E N Y A L A A
ClpXP
>100 μM
>100 μM
Relative Km
mutants D D A A A A A D A D D D
C.
ClpXP
GFP-ssrA
GFP-D2A6DLAA
Loss of GFP-ssrA (min-1[ClpX6]-1)
fluorescence
1 μM
2 μM
time (min)
[GFP] (μM)
D.
A A N D E N Y A L A A
ClpAP
Relative Km
mutants D D A A A A A D A D D D
E.
ClpAP
no peptide
ssrA peptide
ssrA CONH2 peptide
Loss of GFP-ssrA (min-1[ClpX6]-1)
fluorescence
1 μM
2 μM
time (min)
[GFP-ssrA] (μM)

Figure 2.1. Degradation of GFP-ssrA variants by ClpXP and ClpAP.

(A) SsrA-tag sequence and identity of single residue substitutions.

(B) Relative K_m's for ClpXP degradation of GFP-ssrA mutants. Rates of ClpXP-mediated degradation of GFP-ssrA variants, determined by the loss of native fluorescence, were determined at different substrate concentrations (see Methods). and fit to a Michaelis-Menten model. The K_m values plotted were normalized by dividing by K_m for ClpXP degradation of wild-type GFP-ssrA (1.5 μM). V_{max} values for mutants 1-9 were within 2-fold of the wild-type value (1.2 min^{-1} $ClpX_6^{-1}$) except for Y7A which had a V_{max} of 0.45 min^{-1} $ClpX_6^{-1}$).

(C) Michaelis-Menten Plots of ClpXP Degradation of GFP-ssrA and GFP-D_2A_5DLAA. The solid lines are fits to the Michaelis-Menten equation for GFP-ssrA (K_m = 1.5 μM, V_{max} = 1.2 min^{-1}) and GFP-D_2A_5DLAA (K_m = 10.1 μM, V_{max} = 0.8 min^{-1}). The decrease in V_{max} for the consensus mutant is probably caused by the decreased V_{max} of the Y9A substitution. The inset shows the change in fluorescence at 511 nm of 1 μM GFP-ssrA and 2 μM GFP-ssrA following incubation with ClpXP.

(D) Relative K_m's for ClpAP degradation of GFP-ssrA mutants. K_m's were normalized by dividing by the K_m value (1.5 μM) for ClpAP degradation of wild-type GFP-ssrA. See legend to panel B for other details.

(E) Inhibition of ClpAP degradation of GFP-ssrA by ssrA peptides. Michaelis-Menten plots for ClpAP degradation of GFP-ssrA in the absence of peptide (K_m = 1.5 ± 0.4 μM, V_{max} = 4.9 ± 0.3 μM/min^{-1}), or presence of the wild-type ssrA peptide (K_m apparent = 10.4 ± 1.6 μM, V_{max} = 5.1 ± 0.4 μM/min^{-1}, K_I = 16.9 μM), or the carboxamide ssrA peptide (K_m apparent = 10.7 ± 1.2 μM, V_{max} = 4.9 ± 0.3 μM/min^{-1}, K_I = 16.4 μM). K_I values were calculated from K_m apparent = [1 + ([I]/ K_I)]* K_m. The inset shows the change in fluorescence at 511 nm of 1 μM GFP-ssrA and 2 μM GFP-ssrA following incubation with ClpAP.

μM). These results show that ClpA recognizes information in both the N-terminal and C-terminal regions of the ssrA tag.

Because mutation of the C-terminal alanine of the ssrA tag to aspartic acid had no effect on ClpAP degradation, we suspected that the free α-carboxyl group—a unique chemical signature of the C-terminal residue—might also be dispensable. To investigate this question, we compared the ability of peptides with either a normal α-carboxyl group (ssrA peptide) or a terminal carboxamide group (ssrA-$CONH_2$) to inhibit degradation of GFP-ssrA by ClpAP. As shown in Fig. 2.1E, the ssrA-$CONH_2$ peptide (K_i = 16.9 μM) was as effective as the ssrA peptide (K_i = 16.4 μM) in inhibiting degradation of GFP-ssrA by ClpAP. These results suggest that ClpA may be able to recognize an ssrA-like signal in any exposed region of a protein without restriction to the C-terminal end. Previous studies have shown that the α-carboxyl group is an important determinant of ClpX recognition of the ssrA tag, with the ssrA-$CONH_2$ peptide being 10-fold less effective as an inhibitor than the normal ssrA peptide (Kim et al. 2000).

SsrA-tag•SspB recognition.

SspB binds to ssrA-tagged proteins and enhances recognition of these proteins by ClpX. Previous studies showed that SspB binds specifically to the tag, that the N3A tag mutation abrogates this binding, and that deletion of the last three amino acids from the tag does not prevent binding (Levchenko et al. 2000). To define further the interaction between SspB and the ssrA tag, we synthesized an immobilized peptide library in which each residue of the ssrA peptide was individually changed to each of the other 19 amino acids, while the rest of the sequence remained unchanged. These peptides, which contained two additional C-terminal alanines, were covalently attached via their C-termini to a cellulose filter by a

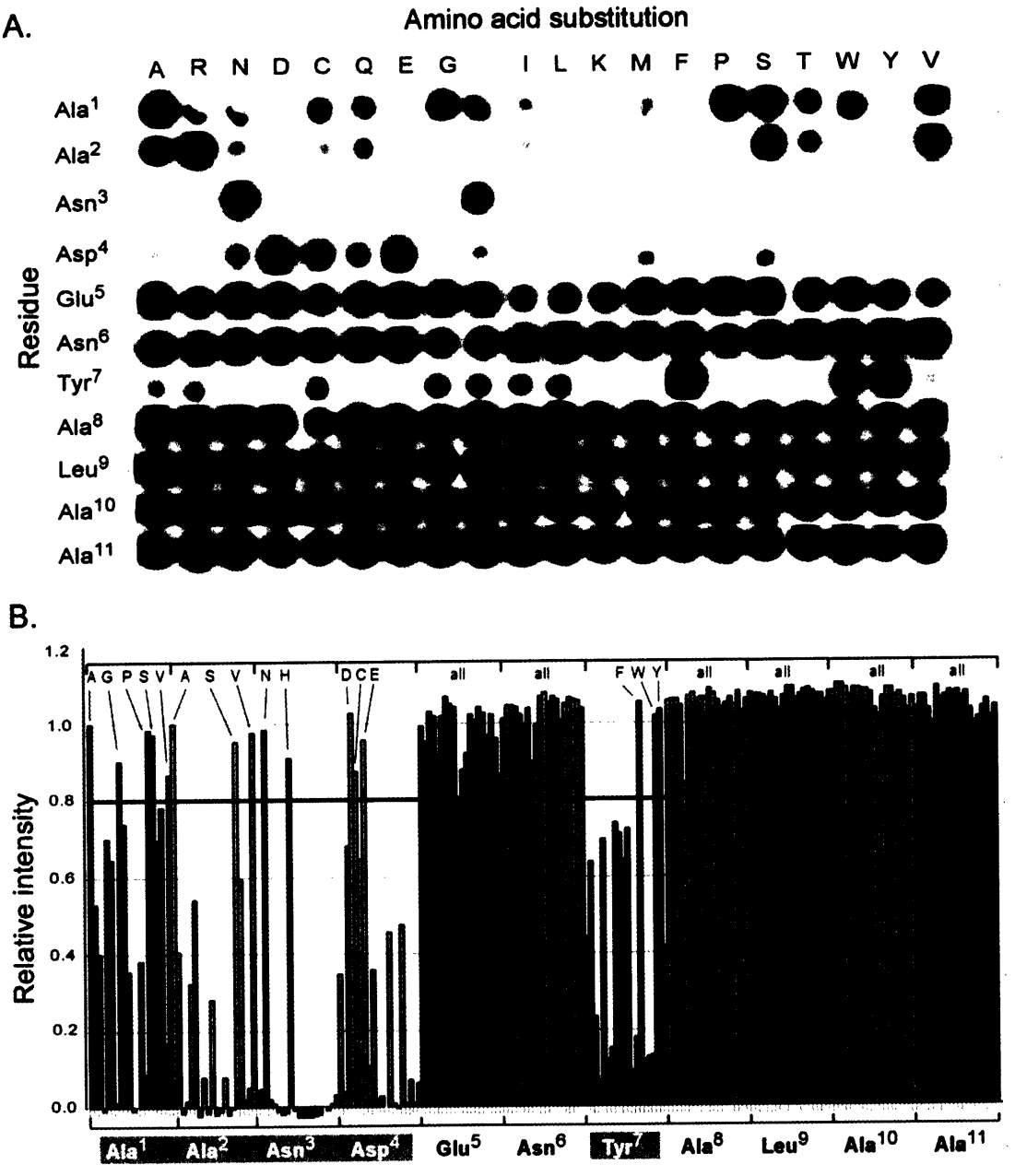

A.
Amino acid substitution
A R N D C Q E G I L K M F P S T W Y V
Residue
Ala1
Ala2
Asn3
Asp4
Glu5
Asn6
Tyr7
Ala8
Leu9
Ala10
Ala11
B.
Relative intensity
1.2
1.0
0.8
0.6
0.4
0.2
0.0
A G P S V A S V N H D C E all all F W Y all all all all
Ala1 Ala2 Asn3 Asp4 Glu5 Asn6 Tyr7 Ala8 Leu9 Ala10 Ala11

Figure 2.2. Effects of ssrA-peptide mutations on SspB recognition.

(A) A library consisting of 220 ssrA peptide variants was used to assay SspB binding via an "indirect" Western. The filter containing covalently bound peptides was first incubated in 10 μg/ml SspB and bound SspB was detected with anti-SspB antibody followed by HRP-conjugated goat anti-rabbit IgG antibody and the ECL substrate.

(B) The filter in (A) was digitally scanned and the number of pixels in each spot was quantified using ImageQuant. These values are presented relative to the intensity of the wild-type ssrA peptide. Substitutions that show 80% or more of wild-type binding are indicated above the graph.

polyethylene glycol linker. The filter contained 220 "spots", with each spot corresponding to one peptide sequence.

Interaction with the peptides was measured by incubating the filter with SspB, and subsequently detecting bound SspB with anti-SspB antibody (Fig. 2.2A). Inspection of the filter showed that SspB bound poorly to many of the peptides with substitutions at tag positions 1, 2, 3, 4 and 7. At position 3, for example, only peptides with Asn or His were efficiently bound. In contrast, at tag positions 5, 6, 8, 9, 10 and 11, SspB had no significant sequence preferences. Fig. 2.2B quantifies the efficiency of the SspB•peptide interactions. Using an arbitrary cut-off value of 80% of wild-type binding produced the consensus: $[AGPSV]^1$-$[ASV]^2$-$[NH]^3$-$[DCE]^4$-X^5-X^6-$[FWY]^7$ for SspB recognition. These results suggest that SspB and ClpX interact with discrete sets of residues in the ssrA tag, whereas SspB and ClpA interact with some of the same residues.

Requirement for dual recognition of the ssrA tag by SspB and ClpX.

Previous studies established that SspB decreases K_m for ClpXP degradation of GFP-ssrA from 1.5 μM to less than 0.3 μM (Levchenko et al. 2000). In principle, binding of SspB to the ssrA tag might be sufficient to target a tagged protein to ClpX without requiring independent recognition of the tag by ClpX. To test this possibility, we monitored degradation in the presence of SspB of the three GFP-ssrA mutants defective in ClpX recognition (L9A, A10D and A11D). Even with SspB, the GFP-ssrA (A10D) and GFP-ssrA (A11D) proteins remained refractory to ClpXP degradation, indicating that binding by SspB does not bypass the requirements for ClpX recognition of these two residues (Fig. 2.3A and data not shown). SspB did, however, enhance recognition of the GFP-ssrA (L9A) mutant by ClpX. In the presence of 0.24 μM SspB, K_m for degradation of GFP-ssrA (L9A) was reduced from 6.2 μM to less than 0.3 μM (Fig. 2.3B). Thus, SspB can compensate for decreased interactions with

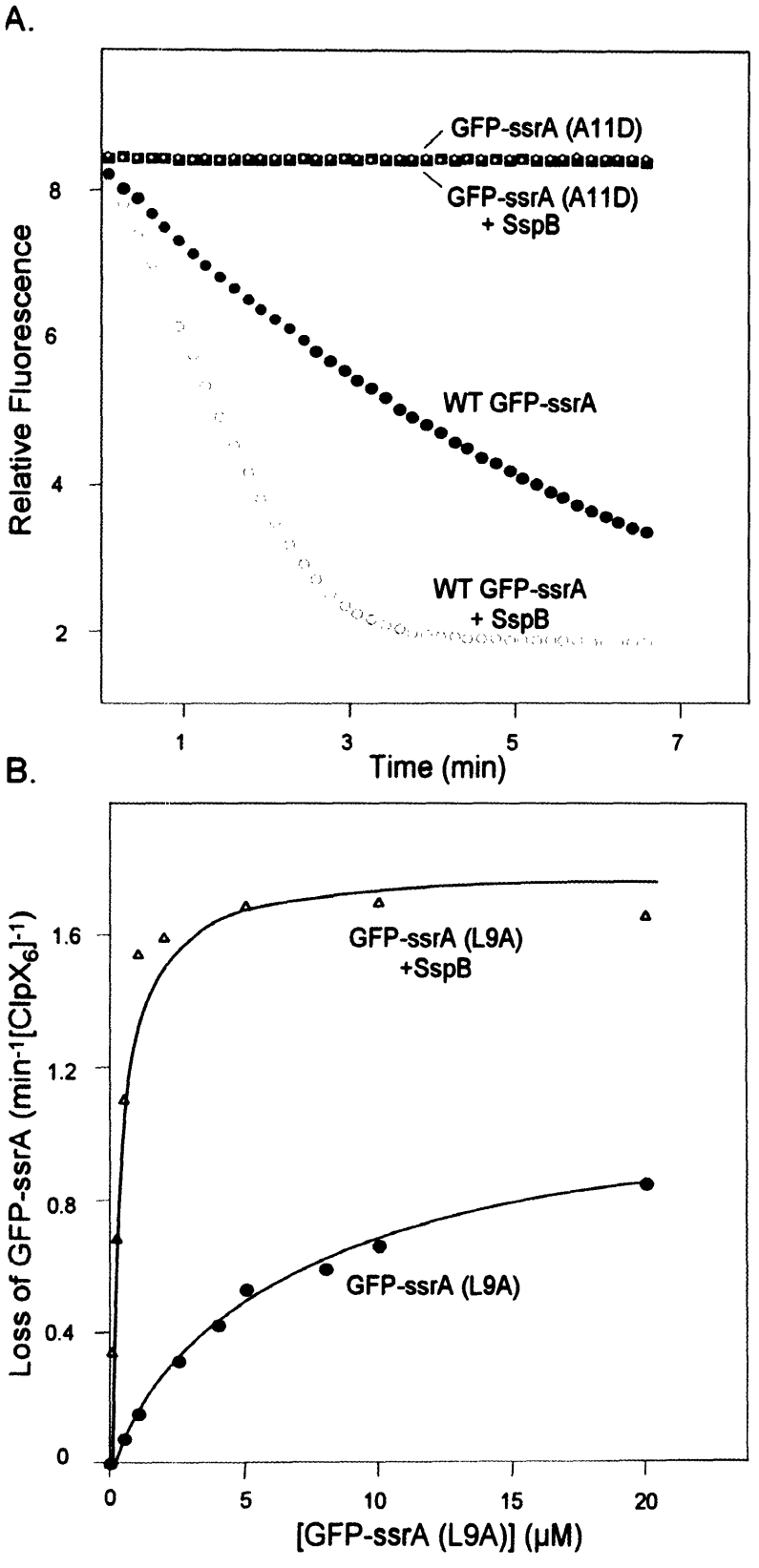

A.
GFP-ssrA (A11D)
GFP-ssrA (A11D)
+ SspB
WT GFP-ssrA
WT GFP-ssrA
+ SspB
Relative Fluorescence
8
6
4
2
1
3
5
7
Time (min)
B.
GFP-ssrA (L9A)
+SspB
GFP-ssrA (L9A)
Loss of GFP-ssrA (min-1[ClpX6]-1)
1.6
1.2
0.8
0.4
0
0
5
10
15
20
[GFP-ssrA (L9A)] (μM)

Figure 2.3.

(A) Degradation of GFP-ssrA (A11D) in the presence of SspB.

ClpXP degradation, assayed by loss of fluorescence at 511 nm of 1 μM GFP-ssrA with or without SspB and 1 μM GFP-ssrA (A11D) with or without SspB. When present, the SspB concentration was 1 μM.

(B) Degradation of L9A in the presence of SspB.

Michaelis-Menten plots for ClpXP degradation of GFP-ssrA (L9A) in the absence (K_m = 6.2 μM, V_{max} = 1.1 min^{-1}) or presence of saturating amounts of SspB ($K_m \leq 0.34$ μM, V_{max} = 1.8 min^{-1}). The K_m represents an upper limit due to the relatively high enzyme concentration (0.3 μM $ClpX_6$) used in the experiment. The solid lines are fits to the Michaelis-Menten equation.

ClpX caused by this mutation. However, the GFP-ssrA (A10D) and GFP-ssrA (A11D) results clearly establish that SspB-regulated degradation of ssrA-tagged proteins depends on both sets of binding determinants, those for ClpX and those for SspB.

SspB inhibits degradation of GFP-ssrA by ClpAP.

Because SspB and ClpA both interact with Ala^1 and Ala^2 in the ssrA tag, it seemed likely that their binding would be mutually exclusive, and thus that SspB could inhibit ClpAP degradation of ssrA-tagged proteins. The results shown in Fig. 2.4 confirm this prediction. ClpAP degradation of GFP-ssrA was completely inhibited in the presence of a two-fold excess of SspB. To ensure that SspB inhibits ClpAP degradation of GFP-ssrA by binding to the ssrA tag, we measured ClpAP degradation of GFP-ssrA (N3A). This mutation prevents binding of SspB to ssrA-tagged GFP (Fig. 2.2 and Levchenko et al. 2000) but does not affect ClpA recognition (Fig. 2.1D). SspB did not inhibit GFP-ssrA (N3A) degradation by ClpAP (Fig. 2.4), indicating that specific interaction of SspB with the ssrA tag is required to inhibit ClpAP degradation of the tagged protein. Thus, SspB binds specifically to the ssrA tag and appears to mask sequence elements important for ClpA interactions.

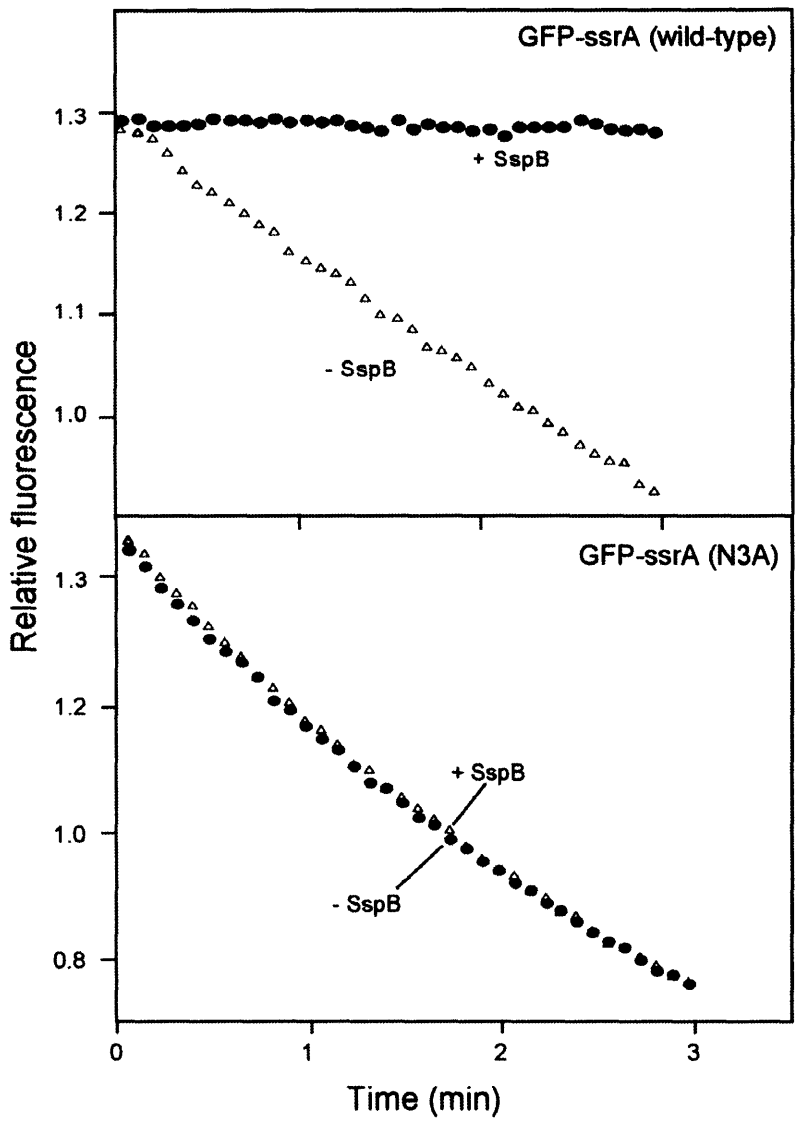


Figure 2.4. SspB inhibits degradation by ClpAP.

ClpAP degradation of 1 μM GFP-ssrA or GFP-ssrA (N3A), assayed by loss of fluorescence at 511 nm, without SspB or with SspB (2 μM).

Discussion

Binding determinants for ClpX and ClpA in the ssrA tag

In *E. coli*, addition of the ssrA degradation tag to a protein is a signal to destroy the resulting polypeptide, and ssrA-tagged proteins are degraded by ClpXP, ClpAP, FtsH (HflB), and Tsp (Prc) (Keiler and Sauer 1996; Gottesman et al. 1998; Herman et al. 1998). The Clp proteases are cytoplasmic, FtsH is a membrane protease, and Tsp is a periplasmic protease, ensuring that tagged proteins are degraded in all cellular compartments. In addition, SspB binds ssrA-tagged proteins in the cytoplasm and enhances their binding to and degradation by ClpXP (Levchenko et al. 2000). Thus, the 11-residue ssrA tag must encode sufficient information to mediate at least five sets of protein-protein interactions. Here, we dissected the sequence elements within the tag that are recognized by SspB, ClpX, and ClpA, the three proteins principally responsible for degradation of ssrA-tagged proteins in the cytoplasm. Our results show that the ssrA tag contains contiguous binding sites for ClpX and SspB but overlapping binding sites for ClpA and SspB (Fig. 2.5A).

The ClpX-binding determinants in the ssrA tag are highly localized, composed of the α-carboxyl group and C-terminal residues, Leu^{9}-Ala^{10}-Ala^{11}. Within this set, however, aspartic-acid substitutions at Ala^{10} or Ala^{11} completely blocked substrate recognition by ClpX and were far more deleterious than substitutions at Leu^{9} or the α-carboxyl group. Interestingly, Tsp recognizes ssrA-tagged polypeptides (Beebe et al. 2000) and non ssrA-tagged substrates that end with Leu-Ala-Ala (Keiler and Sauer 1996), indicating that this protease interacts with the same portion of the ssrA tag as ClpX. Although the ClpX determinants are highly localized at the C-terminal end of the ssrA tag, it is important to note that GFP-G_8LAA, which has the terminal Leu-Ala-Ala sequence, was not degraded by ClpXP. These data suggest that the sequence context or structure of a peptide containing this terminal tripeptide can influence ClpX interactions.

About half of the known ClpX substrates are similar to ssrA-tagged proteins in having non-polar side chains at the penultimate and C-terminal residues (Levchenko et al. 1997b). This group includes MuA (Ala-Ile), Mu repressor (Ala-Val), Mu repressor vir 3061 (Val-Leu), and CtrA (Ala-Ala). In several cases, these non-polar residues have been implicated in ClpX recognition (Laachouch et al. 1996; Domian et al. 1997; Levchenko et al. 1997b). It seems likely that ClpX uses the same substrate-binding site to interact with each of these substrates and with the ssrA tag. In contrast, other ClpX substrates — λO (Gonciarz-Swiatek et al. 1999), UmuD' (Gonzalez et al. 2000), TrfA (Konieczny and Helinski 1997), Phd (Lehnherr and Yarmolinsky 1995) and σ^s (Zhou et al. 2001) — lack non-polar residues at their C-termini. Furthermore, where determined, the sequences responsible for protease targeting in these proteins have been localized to regions other than the C-terminus. Thus, it is an attractive model that these proteins are recognized by ClpX using a different binding surface than the one that recognizes the ssrA tag.

Rules governing substrate recognition by ClpA are currently poorly defined. Our mutational analysis reveals that the most important residues of the ssrA tag for recognition by ClpA were Ala^1, Ala^2, Ala^8, Leu^9 and Ala^{10}, with the substitutions A1D, A8D and L9D being especially deleterious. Thus, the ClpA recognition determinants, like those of ClpX, involve aliphatic side chains. Unlike the ClpX determinants, however, those for ClpA are not highly localized. It is unclear whether the 5-residue spacing between the Ala^1-Ala^2 and Ala^8-Leu^9-Ala^{10} determinants is important for ClpA recognition. Surprisingly, GFP-D_2A_5DLAA was found to be efficiently degraded by ClpAP *in vitro* ($K_m \approx 2$ μM, unpublished observations). The tag of this substrate does not contain several important ClpA-recognition determinants nor does it contain a $\Phi\Phi X_5\Phi\Phi\Phi$ motif (where Φ represents an aliphatic side chain). This tag does, however, contain $\Phi\Phi X_4\Phi\Phi\Phi$, $\Phi\Phi X_3\Phi\Phi\Phi$ and $\Phi\Phi X_2\Phi\Phi\Phi$ motifs, suggesting that ClpA might recognize short clusters of aliphatic residues with variations in spacing.

ClpX and ClpA are related proteins that both recognize the ssrA tag. Thus, it was a reasonable hypothesis that they might share homologous substrate binding pockets responsible for this common substrate recognition. However, we find that these ATPases achieve common recognition of the ssrA tag by interacting with different sequences in the peptide (Fig. 2.5A). This finding clearly favors the idea that the ssrA tags are recognized by these two proteins using substrate-binding pockets with substantially different recognition characteristics. Consistent with this conclusion, ClpX and ClpA generally recognize distinct proteins.

SspB is a bifunctional regulator of substrate recognition.

SspB exhibits strong preferences for specific side chains at positions 1, 2, 3, 4, and 7 of the ssrA tag. These SspB-binding determinants are adjacent to those recognized by ClpX, allowing both proteins to bind to the same ssrA tag. Mutual binding, in this instance, is required for SspB to stimulate ClpXP degradation of ssrA-tagged substrates. Disruption of either SspB or ClpX recognition of the ssrA tag abolishes efficient degradation of ssrA-tagged substrates by ClpXP (see Fig. 2.3 and Levchenko et al. 2000). Consistent with this substrate docking mechanism, ClpX, SspB, and an ssrA-tagged substrate form stable terinary complexes (Levchenko et al. 2000). In contrast, the SspB-binding determinants in the ssrA tag overlap those for ClpA recognition, and SspB, as a consequence, inhibits ClpAP degradation of ssrA-tagged substrates. Hence, SspB binding to the ssrA-tagged substrates enhances their degradation by ClpXP but inhibits proteolysis by ClpAP. SspB's ability to divert ssrA-tagged substrates from ClpAP to ClpXP helps explain the observation that both proteases degrade ssrA-tagged proteins similarly *in vitro,* whereas these substrates are preferentially degraded by ClpXP *in vivo* (Gottesman et al. 1998; Levchenko et al. 2000).

Is SspB-mediated channeling of ssrA-tagged substrates from ClpAP to ClpXP biologically important? The answer to this question is uncertain, but the different activities of

the two proteases towards certain substrates provides an opportunity for speculation. For example, ClpAP but not ClpXP degrades unfolded proteins without targeting signals (Katayama et al. 1988; Hoskins et al. 2000a), an activity that is probably most important during heat shock or other types of environmental stress. Up-regulation of SspB in response to stress could redirect ssrA-tagged substrates to ClpXP, leaving ClpAP free to degrade unfolded substrates.

Conservation of ClpX- and SspB-recognition modules within the ssrA tag

The C-terminal tripeptide of the ssrA tag from a variety of bacterial species is highly conserved (LAA or VAA; Fig. 2.5B), consistent with the observation that ClpX and Tsp orthologs, which are likely to recognize these positions, are present in these bacteria. SspB orthologs are only found in the gamma- and beta-proteobacteria (Levchenko et al. 2000). Alignment of the ssrA tags from these bacteria (Fig. 2.5B) reveals a consensus for the first seven tag residues, $[A]^1$-$[A]^2$-$[N]^3$-$[DE]^4$-$[SDE]^5$-$[TNRQ]^6$-$[YF]^7$, that is a subset of the *E. coli* SspB consensus, $[AGPSV]^1$-$[ASV]^2$-$[NH]^3$-$[DCE]^4$-X^5-X^6-$[FWY]^7$, determined here. The N-terminal portions of ssrA-tag sequences from other bacterial families are still highly conserved (Fig. 2.5B), although clearly distinct from the sequence bound by SspB. These observations suggest either that these bacteria contain an SspB-like regulator or that these regions are conserved because they mediate interactions with other proteases.

A. Stimulatory SspB•ClpX complex

EN ALAA-CO_2^-

Competition between ClpA and SspB binding

EN ALAA-CO_2^-

AANDENYALAA-CO_2^-

Recognition determinants for
- ClpX
- ClpA
- SspB

B.

Gamma Proteobacteria	SsrA tag
E. coli	EN ALAA
V. cholerae	EN ALAA
Y. pestis	EN ALAA
S. enteritidis	ET ALAA
X. fastidiosa	DN AVAA
P. aeruginosa	DN ALAA
H. influenzae	EQ ALAA
Beta Proteobacteria	
N. gonorrhoeae	ET ALAA
N. meningitidis	ET ALAA
B. pertussis	ER ALAA
Firmicutes, Bacillus/Staphylococcus	
B. subtilis	T SFNQN---VALAA
B. anthracis	Q-------LSLAA
B. megaterium	------ALAA
S. aureus	------AVAA
S. epidermidis	D------AVAA
Cyanobaceria	
Synechocystis PCC6803	S K---VALAA
P. marinus	K S QT- PVAA
Cyanobacterium PCC6307	R QA- PVAA
Synechococcus WH8102	R HA- PVAA
N. punctiforme	N A KD- LVAA
Avabaena PCC7120	K A KD- LVAA

Figure 2.5.

(A) Recognition determinants within the ssrA tag for ClpX, ClpA and SspB. Recognition determinants for ClpX are highlighted in black, those for ClpA in dark gray, and those for SspB in light gray.

(B) SsrA-degradation tags from different bacteria. The conserved SspB binding determinants in the gamma and beta proteobacteria are highlighted in light gray. Shown are the predicted ssrA tag sequences from representative members of various families of bacteria. The conserved residues in the N-terminal regions of the ssrA tag in the other families are highlighted in dark gray. All sequenced gamma and beta proteobacteria have a predicted ssrA tag sequence that contains an acceptable SspB binding site with the exception of *Buchnera sp.*, strain APS (tag sequence: (A)ANNKQNYALAA). Interestingly, this bacterium does not have a detectable ortholog of SspB. Of the bacteria listed, the following appear to have a ClpA ortholog, in addition to a ClpX: *E. coli, V. cholerae, X. fastidiosa,* and *P. aeruginosa.*

Acknowledgements

We thank members of the Sauer and Baker labs for advice and help. Supported by NIH grant AI-16892 and the Howard Hughes Medical Institute. T. A. Baker is an employee of the Howard Hughes Medical Institute.

Experimental Procedures

Materials: ClpX (Levchenko et al. 1997a), ClpP (Kim et al. 2000), SspB (Levchenko et al. 2000), ClpA (Maurizi et al. 1994) and GFP-ssrA (Yakhnin et al. 1998) were purified as described. Polyclonal anti-SspB antibodies were prepared by Covance (Denver, PA), using SspB purified in our laboratory. PD buffer (pH 7.6) contains 25 mM HEPES-KOH, 5 mM $MgCl_2$, 5 mM KCl, 15 mM NaCl, 0.032% (v/v) NP-40, 10% (v/v) glycerol. HO buffer (pH 7.5) contains 25 mM HEPES-KOH, 20 mM $MgCl_2$, 300 mM NaCl, 10% (v/v) glycerol and 0.5 mM DTT.

GFP mutants: A gene encoding GFP-ssrA with S6G and S72A mutations in the GFP coding sequence (GFPmut3-ssrA) (Andersen et al. 1998), a gift of A.J. Anderson (The Technical University of Denmark), was cloned into the Not I site of pACYC184 to create pMS30. Mutant ssrA tags were introduced by ligating the StuI and HindIII cleaved backbone fragment of pMS30 to synthetic oligonucleotide cassettes. DNA sequences were determined for all GFP-ssrA variants to confirm the expected sequence. The molecular weights of GFP-ssrA (A10D) and GFP-ssrA (A11D) were confirmed by mass spectrometry.

Degradation assays: $ClpX_6$ (0.3 μM), $ClpP_{14}$ (0.8 μM), ATP (4 mM), and an ATP regeneration system (50 μg/ml creatine kinase and 2.5 mM creatine phosphate) were mixed in PD buffer and incubated for 2 min at 30 °C. GFP-ssrA or variants were then added and the mixture was transferred to a 50 μl cuvette, and fluorescence readings were begun within 10 sec. In some reactions, SspB was added, in concentrations indicated in the figure legends (as monomer equivalents), following the 2 min incubation at 30 °C but prior to addition of substrate. Changes in GFP fluorescence (excitation 467 nm; emission 511 nm) were monitored in a

Fluoromax-2 instrument (ISA Jobin Yvon-Spex). Degradation of GFP-ssrA or variants by ClpAP was performed as above except using $ClpA_6$ (0.05 μM) and $ClpP_{14}$ (0.1 μM) in HO buffer. Reaction solution conditions for ClpXP and ClpAP were different in order to optimize the activity observed for each enzyme.

Peptide•SspB binding: A cellulose filter containing 220 synthetic ssrA peptide variants was prepared by the MIT Biopolymers facility using an Abimed instrument. Each peptide contained two additional alanines, C-terminal to the end of the ssrA sequence, and was covalently attached to the filter via a polyethylene glycol linker. The filter was blocked for 3 hours in TBST (50 mM Tris [pH 7.5], 125 mM NaCl, 0.1% Tween-20) plus 10% milk; incubated with 10 μg/ml SspB in TBST plus 0.1% milk; washed three times in TBST; incubated with polyclonal rabbit anti-SspB antibody for 1 hour; washed three times in TBST; and incubated for 30 min with secondary goat anti-rabbit IgG HRP conjugated antibody (Amersham Life Sciences). Three final washes with TBST were performed; the filter was incubated with ECL substrate (NEN); and binding was visualized on film. Attempts to probe ClpX or ClpA binding to the peptide filter were unsuccessful.

CHAPTER THREE:
Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals

---

This chapter was previously published as Flynn, J.M., S.B. Neher, Y.I. Kim, R.T. Sauer, and T.A. Baker. *Molecular Cell* **11**: 671-83 (2003). S.B. Neher contributed data to Figure 3.2a and showed that the LexA autocleavage products are ClpXP substrates. Y.I. Kim cloned the $ClpP^{trap}$. R.T. Sauer and T.A. Baker assisted in preparation of the manuscript.

Abstract

ClpXP is a protease involved in DNA-damage repair, stationary-phase gene expression, and ssrA-mediated protein quality control. To date, however, only a handful of ClpXP substrates have been identified. Using a tagged and inactive variant of ClpP, substrates of *E. coli* ClpXP were trapped *in vivo*, purified, and identified by mass spectrometry. The more than 50 trapped proteins include transcription factors, metabolic enzymes, and proteins involved in the starvation and oxidative stress responses. Analysis of the sequences of the trapped proteins revealed five recurring motifs: two located at the C-terminus of proteins, and three N-terminal motifs. Deletion analysis, fusion proteins, and point mutations established that sequences from each motif class targeted proteins for degradation by ClpXP. These results represent the first description of general rules governing substrate recognition by a AAA+-family ATPase and suggest strategies for regulation of protein degradation.

Introduction

Protein degradation is an essential component of biological regulation and protein quality control in organisms ranging from bacteria to humans. Many cytoplasmic proteases are large multi-subunit complexes in which the proteolytic active sites are sequestered within an internal chamber. Access to this chamber is controlled by axial pores that exclude native proteins and all but the smallest peptides (for review, see Lupas et al. 1997). These multimeric proteases form complexes with AAA+ ATPases, which denature and translocate substrates into the proteolytic chamber for degradation (for review, see Ogura and Wilkinson 2001). The ClpXP, ClpAP, HslUV (ClpYQ), HflB (FtsH) and Lon proteases of bacteria share this basic mechanism with the proteasomes of eukaryotic organisms (for review, see Schirmer et al. 1996). Identifying the proteolytic targets of specific proteases is critical to any general understanding of their diverse cellular functions and provides a way to decipher the rules by which these enzymes recognize substrates.

E. coli ClpXP is an ATP-dependent intracellular protease. The ClpX component is a hexameric AAA+ ATPase responsible for substrate recognition, unfolding, and translocation into ClpP (Wojtkowiak et al. 1993; Wawrzynow et al. 1995; Weber-Ban et al. 1999; Kim et al. 2000). ClpX can also act independently to dismantle multimers and remodel proteins (Levchenko et al. 1995). ClpP is a 14-subunit serine peptidase (Maurizi et al. 1990a). It has a barrel-like structure comprised of two heptameric rings. Face-to-face stacking of these rings sequesters the active sites within the proteolytic chamber (Wang et al. 1997). One or two ClpX hexamers bind to $ClpP_{14}$ to form the ClpXP protease (Grimaud et al. 1998). ClpP also combines with hexamers of the ClpA ATPase to form ClpAP (Katayama et al. 1988). ClpX and ClpA generally confer distinct substrate specificities to their respective protease complexes although these enzymes do recognize some common substrates (for review, see Gottesman, 1996; Gottesman et al. 1998).

ClpX and ClpP orthologs are found in most bacteria, mitochondria, and chloroplasts. In *E. coli*, *clpP*-defective cells show delayed recovery both from stationary phase and following a shift to nutrient poor media (Damerau and St John 1993). Proteolysis by ClpXP is involved in the

development of competence and in sporulation in *Bacillus subtilis* and is required for viability and cell-cycle progression in *Caulobacter crescentus* (Jenal and Fuchs 1998; Msadek et al. 1998). ClpP is also important for the virulence of bacterial pathogens including *Yersinia enterocolitica*, *Streptococcus pneumoniae*, *Salmonella typhimurium*, and *Listeria monocytogenes* (for review, see Porankiewicz et al. 1999).

Despite the diverse physiological roles of ClpXP, only a few substrates have been identified. *E. coli* ClpX was originally discovered as a component required for ClpP-dependent degradation of the λO phage replication protein (Gottesman et al. 1993). Since then, four additional phage or plasmid proteins (Mu repressor, MuA transposase, RK2 replication protein TrfA, and the P1 antidote protein PhD) and three *E. coli* proteins (the stationary-phase sigma factor σ^S, the SOS protein UmuD', and a type I restriction-modification subunit HsdR) have been identified as ClpXP substrates (see Gottesman 1996 and references therein; also see Frank et al. 1996; Konieczny and Helinski 1997; Makovets et al. 1998). ClpXP also degrades proteins modified by addition of the ssrA tag, an 11-residue sequence added cotranslationally to the C-terminus of nascent polypeptides on stalled ribosomes ((Keiler et al. 1996; Gottesman et al. 1998).

ClpX interacts with peptide sequences—referred to as recognition signals—at the C-termini of the ssrA tag and MuA (Levchenko et al. 1997b; Gottesman et al. 1998). In contrast, signals near the N-terminus of λO appear most important for ClpX recognition (Gonciarz-Swiatek et al. 1999). In addition to these examples of direct recognition, auxiliary proteins are implicated in targeting some substrates to ClpXP; UmuD confers instability to UmuD' (Gonzalez et al. 2000) and RssB targets σ^s to ClpXP (Muffler et al. 1996). Although progress is being made in understanding how ClpX recognizes some members of a small group of substrates, general rules governing substrate recognition have yet to emerge.

Here, we report the identification of more than 50 *E. coli* proteins that are trapped in a ClpX-dependent fashion within an active-site mutant of ClpP. Analysis of these ClpXP substrates provides a more comprehensive understanding of the cellular roles of this protease and reveals five distinct

classes of ClpX-recognition motifs. This study provides the first general description of the sequence rules that mediate substrate recognition by an energy-dependent intracellular protease and establishes a foundation for understanding how degradation may be regulated.

Results

Protein trapping by ClpXP in vivo.

To identify new substrates, we took advantage of the ability of inactivated ClpP to accept and retain proteins translocated into its chamber by the ClpX ATPase (Kim et al. 2000). A $ClpP^{trap}$ variant was constructed containing an active-site mutation (S97A) as well as a C-terminal tandem-affinity tag (Myc_3-TEV-His_6). Proteins translocated into the proteolytic chamber of $ClpXP^{trap}$ *in vitro* were not degraded and were only released slowly (Kim et al. 2000; Singh et al. 200; data not shown). To test whether $ClpP^{trap}$ also captured substrates *in vivo*, it was co-expressed in *E. coli* with GFP-ssrA, a model ClpXP substrate. GFP-ssrA co-purified with $ClpP^{trap}$ during affinity chromatography, confirming that trapping occurred *in vivo* (data not shown). Cellular trapping of GFP-ssrA was prevented by an ssrA-tag mutation (C-terminal A→D, data not shown) that prevents ClpXP degradation *in vitro* (Flynn et al. 2001) indicating that trapping requires the same ClpX-substrate interactions needed for degradation.

To determine if capture by $ClpP^{trap}$ depended on the ClpX or ClpA ATPases, experiments were performed in *clpX*$^+$*clpA*$^+$, *clpX*$^+$*clpA*$^-$, *clpX*$^-$*clpA*$^+$ and *clpX*$^-$*clpA*$^-$ strains. To avoid trapping a heterogeneous collection of ssrA-tagged proteins, we deleted the gene encoding SmpB, a protein required for ssrA-tagging (Karzai et al. 1999), from the trapping strains. These strains also carried an insertion in the chromosomal copy of *clpP* and expressed $ClpP^{trap}$ under control of an IPTG-inducible promoter. Proteins that co-purified with $ClpP^{trap}$ in each strain were visualized by staining after electrophoresis on 2D gels (Fig. 3.1).

Approximately 70 proteins co-purified with $ClpP^{trap}$ in the strain expressing both ClpX and ClpA (Fig. 3.1b). A subset of approximately 50 of these proteins were trapped in the strain expressing just ClpX (Fig. 3.1c) whereas about 30 proteins were trapped in the strain expressing just ClpA (Fig 1d). In the absence of ClpX and ClpA, only a handful of polypeptides co-purified with

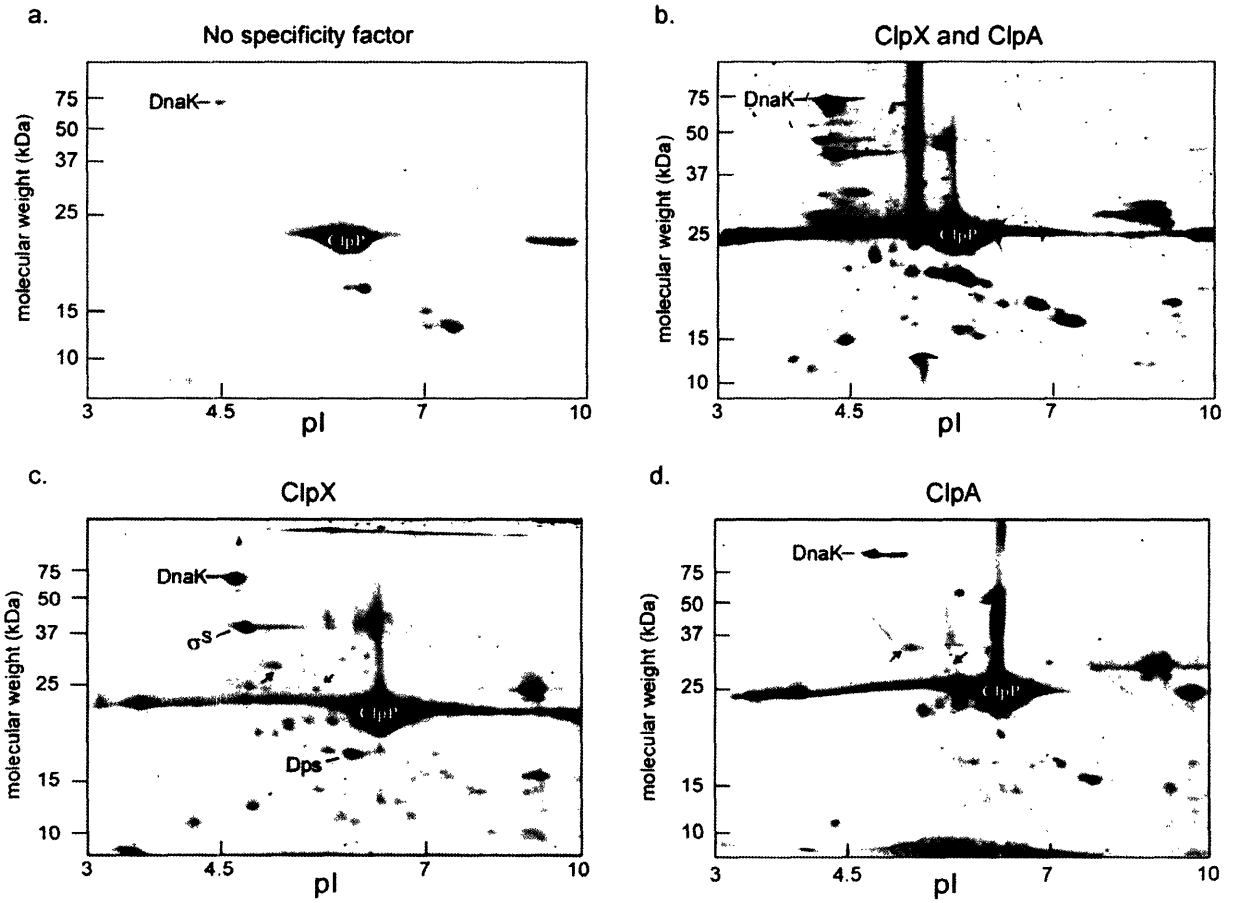


Figure 3.1. 2D-gel analysis of proteins captured by $ClpP^{trap}$.

Panels show proteins captured by $ClpP^{trap}$ in *E. coli* strains JF148 (**a**), JF176 (**b**), JF162 (**c**) and JF172 (**d**). Arrows indicate representative proteins captured by both $ClpXP^{trap}$ and $ClpAP^{trap}$.

ClpP^{trap} (Fig. 3.1a), most of which were shown to be ClpP fragments (data not shown). DnaK also co-purified with ClpP^{trap} in cells lacking both ClpX and ClpA (see Discussion). Because the identities of the vast majority of ClpP^{trap}-captured proteins depended on the presence of ClpX or ClpA, we conclude that these ATPases selectively recognize and translocate proteins into the trap. The proteins captured in a ClpX-dependent or ClpA-dependent fashion are therefore likely to be substrates for degradation by ClpXP or ClpAP. About 10 proteins were present in both the ClpX only and ClpA only samples (Fig. 3.1c & 3.1d), suggesting that these proteins are substrates for both proteases (see Table 3.1). Below, we characterize many of the proteins captured by ClpP^{trap} in a ClpX-dependent manner.

Identification of ClpXP substrates.

To identify cellular proteins captured by ClpXP^{trap}, complexes were isolated from the strain containing ClpX but not ClpA (*clpX*$^{+}$*clpA*$^{-}$) and separated on a 1D gel. Gels slices were excised, digested with trypsin, and analyzed by tandem-mass spectrometry. This procedure identified 60 *E. coli* proteins in addition to ClpP, ClpX, and the TEV protease (Table 3.1). One of the most abundant trapped proteins was σ^S (Fig. 3.1c), the stationary-phase sigma factor that is degraded by ClpXP during exponential growth (Schweder et al. 1996). Proteins captured by ClpXP^{trap} included a wide variety of regulatory proteins and biological catalysts (Table 3.1) including many with suggested roles in stationary phase and oxidative stress responses (see Discussion). Based on annotations, nearly all of these proteins reside in the cytoplasm with ClpXP. One outer membrane protein, OmpA, and one inner membrane protein, RseA, were apparent exceptions (see below and Discussion). Mass spectrometry of the *clpX*$^{-}$*clpA*$^{-}$ sample revealed the presence of peptides from only two of the 60 *E. coli* proteins trapped in the *clpX*$^{+}$*clpA*$^{-}$ strain (see Experimental Procedures), providing further evidence of the importance of ClpX for the observed capture.

Table 3.1. ClpXP^{trap} associated proteins.

Gene	SwissProt accession #	# peptides	C-terminal signal	N-terminal signal	Gene product or function
Transcriptional regulators					
crl	P24251	9	C-M1*	N-M3 ++	Curlin genes regulatory protein
dksA	P18274	6	C-M1	N-M3 ++	DnaK suppressor protein
fnr	P03019	8	C-M1	N-M1 ++	Transcription regulator FNR
iscR	P77484	9	C-M1	N-M2 ++	Iron-sulfur cluster regulator
lexA	P03033	3		N-M2 ++	LexA repressor
rpoS (σ^S)	P13445	110		N-M1 +	RNA polymerase sigma factor σ^S
rsd	P31690	1	C-M2		Regulator of sigma D
rseA	P38106	2			Negative regulator of sigma-E
Translation					
rplE	P02389	2		N-M1 +	50S ribosomal protein L5
rplJ	P02408	57	C-M1	N-M3 +	50S ribosomal protein L10
rplK	P02409	5		N-M1 +	50S ribosomal protein L11
rplN	P02411	17			50S ribosomal protein L14
rplS	P02420	11		N-M1 +	50S ribosomal protein L19
rplU	P02422	2	C-M1	N-M3 +	50S ribosomal protein L21
tufB	P02990	2		N-M3 +	Elongation factor Ef-Tu
Chaperones & degradation					
clpX	P33138	5		N/A	Clp protease ATP-binding subunit
dnaK	P04475	75		N/A	Chaperone Hsp70
gcp	P05852	7	C-M1		O-sialoglycoprotein endopeptidase
groEL	P06139	6		N/A	Chaperone Hsp60
lon	P08177	3			ATP-dependent protease Lon
pepB	P37095	2	C-M1		Aminopeptidase B
Detoxification (protection)					
dps	P27430	70		N-M1 ++	Global regulator protein Dps
katE	P21179	1	C-M1	N-M3 +	Hydroperoxidase II
nrdH	Q47414	2	C-M1	N-M2 +	Glutaredoxin-like protein NrdH
tpx	P37901	4	C-M1	N-M1 +	Thiol peroxidase
Cell division					
ftsZ	P06138	3		N/A	Cell division GTPase
Transposition					
insH	P03837	4		N-M3 ++	IS5 transposase
Cell motility and transport proteins					
cheW	P07365	2	C-M1	N-M1 ++	Chemotaxis protein CheW
cysA	P16676	13	C-M1	N-M1 ++	Sulfate permease A protein
exbB	P18783	5	C-M1		Uptake of enterochelin
gatA	P37187	5		N-M1 ++	Galactitol-specific enzyme IIA
ompA†	P02934	4		N-M2 +	Outer membrane protein 3a
secA	P10408	5		N-M1 +	Protein translocase protein SecA

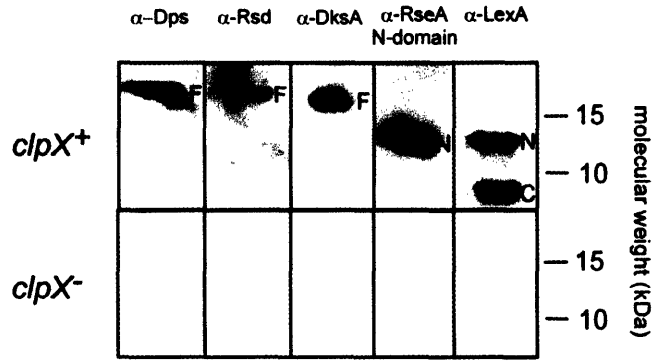
Metabolism & energy production					
aceA†	P05313	12		N-M2 ++	Isocitrate lyase
acnB	P36683	1	C-M1		Aconitase
aldA	P25553	1	C-M1		Aldehyde dehydrogenase
atpD	P00824	6		N-M1 ++	β subunit of F1 ATP synthase
cysD	P21156	2		N-M1 +	Sulfate adenylyltransferase
dadA	P29011	1		N-M2 ++	D-amino acid dehydrogenase
fabB	P14926	1		N-M2 ++	β-ketoacyl-acyl carrier protein synthase I
gapA†	P06977	4		N-M1 ++	Glyceraldehyde 3-phosphate dehydrogenase
gatY	P37192	3		N-M2 ++	Tagatose 1,6-bisphophate aldolase
gatZ	P37191	5		N-M2 ++	Tagatose 6-phosphate kinase
glcB	P37330	2		N-M3 ++	Malate synthase
glpD	P13035	1	C-M1		Glycerol 3-phosphate dehydrogenase
glyA	P00477	2	C-M1	N-M1 +	Glycine hydroxymethyltransferase
iscS	P39171	1		N-M2 ++	Cysteine desulferase
iscU	P77310	4			IscU
lipA	P25845	3		N-M2 ++	Lipoic acid synthetase
lldD	P33232	2	C-M1*		L-Lactate dehydrogenase
moaA	P30745	3		N-M3 +	Molybdopterin biosynthesis, protein A
paaA	P76077	1	C-M2		Phenylacetic acid degradation protein
pncB	P18133	4	C-M2	N-M1 +	Nicotinate phosphoribosyltranferase
ribB	P24199	8	C-M2*		Riboflavin biosynthase
tnaA†	P00913	32		N-M1 +	Tryptophanase
udp	P12758	1		N-M1 +	Uridine phosphorylase
Unknown function					
ybaQ	P77303	1	C-M2*	N-M3 +	
ycbW	P75862	5	C-M1	N-M2 ++	
ydaM	P77302	3	C-M1*		
yebO	P76266	4			
ygaT	P76621	7		++	

Table 3.1.
Proteins are grouped into functional categories based on annotations from the SwissProt database (Bairoch and Apweiler, 2000) and the general literature. For each protein, the gene name, SwissProt accession number, number of peptides identified by MS/MS analysis, and protein name are listed. Proteins with C-terminal sequences similar to those of the ssrA tag (C-M1) or the MuA tag (C-M2) are marked. * indicate proteins whose corresponding C-terminal peptides inhibit ClpXP degradation of GFP-ssrA. † indicate proteins that were also found to be captured by $ClpAP^{trap}$. Proteins whose N-terminal peptides bind to ClpX strongly (++) or moderately (+) are marked. GroEL, FtsZ, ClpX, and DnaK were not tested for binding of their N-termini to ClpX (N/A). The N-termini of the proteins that bind to ClpX are categorized as containing N-motif 1 (N-M1), N-motif 2 (N-M2) or N-motif 3 (N-M3) as defined in Fig. 4.4b. A western blot revealed the presence of Rsd in $ClpP^{trap}$ (see Fig. 3.2), establishing that the identity of trapped proteins can be determined reliably from a single peptide.

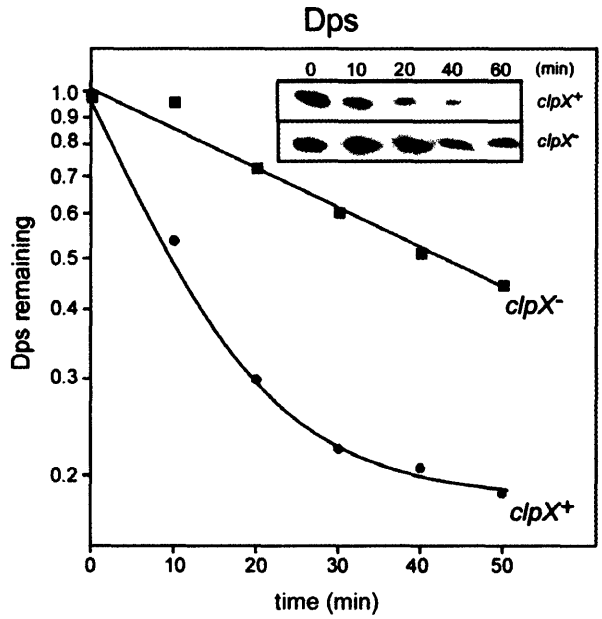
Western blots confirmed ClpX-dependent trapping of five proteins and also established whether full-length proteins or fragments were captured (Fig. 3.2a, upper panel). Antibodies against Dps, Rsd, and DksA reacted with species having molecular weights expected for each full-length protein. In contrast, protein fragments rather than the full-length RseA and LexA co-purified with the $ClpP^{trap}$ (Fig. 3.2a, upper panel). For LexA, the two antibody-reactive bands had the electrophoretic mobility expected for protein fragments generated by RecA-mediated auto-cleavage between Ala^{84} and Gly^{85} (Little et al. 1980). For RseA, the trapped fragment bound antibodies that recognize the protein's N-terminal, cytoplasmic domain. These data strongly suggest that trapping of RseA and LexA depends upon initial cleavage of these proteins by other proteases (see Discussion). None of the five proteins tested were detected in trapped complexes isolated from the *clpX*$^-$*clpA*$^-$ strain (Fig. 3.2a, lower panel) confirming the specificity of trapping.

Degradation experiments support the hypothesis that proteins that co-purify with $ClpXP^{trap}$ are substrates for ClpXP degradation. For example, Dps, a DNA-binding protein induced during starvation (Almiron et al. 1992) and one of the most abundant trapped proteins, had a significantly longer half-life in *clpX*$^-$ than in *clpX*$^+$ cells during outgrowth from stationary phase (Fig. 3.2b) and was efficiently degraded by ClpXP *in vitro* (see below). DksA, the *dnaK* suppressor protein (Kang and Craig 1990), was also stabilized in the *clpX*$^-$ strain, suggesting that ClpXP participates in degradation of this protein *in vivo* (Fig. 3.2b). Note, however, that other proteases must also contribute to the degradation of Dps and DksA *in vivo* because these proteins were still degraded in the *clpX*$^-$ strain (see Discussion). The N-terminal and C-terminal auto-cleavage fragments of LexA were also found to be degraded by ClpXP *in vivo* and *in vitro* (Neher et al. 2003a). Finally, E.L. Mettert and P.J. Kiley (personal communication) demonstrated that another trapped protein, the transcription regulator Fnr, was degraded in a ClpXP-dependent manner *in vivo* when cells were grown aerobically. Hence, σ^S, GFP-ssrA, the LexA N-domain and C-domain, Dps, DksA and Fnr are both captured by $ClpXP^{trap}$ and appear to be substrates for ClpXP degradation. These results demonstrate the effectiveness of

a.


α-Dps
α-Rsd
α-DksA
α-RseA N-domain
α-LexA
clpX+
clpX-
F
N
C
15
10
15
10
molecular weight (kDa)

b.


Dps
0 10 20 40 60 (min)
clpX+
clpX-
Dps remaining
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
clpX-
clpX+
0 10 20 30 40 50
time (min)

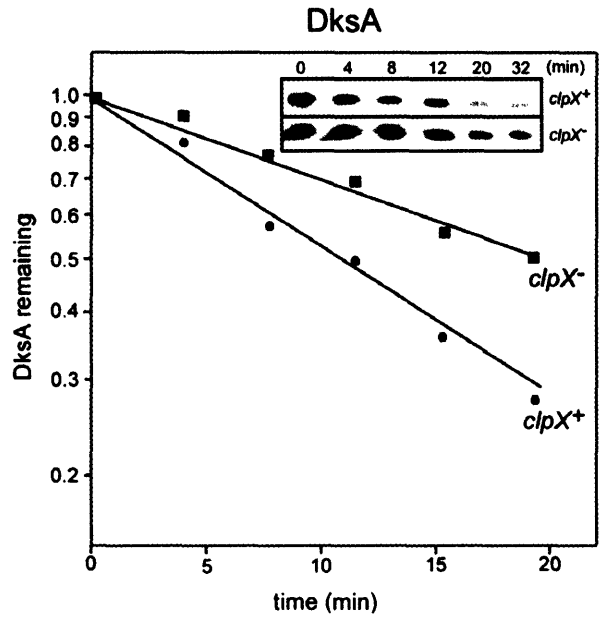

DksA
0 4 8 12 20 32 (min)
clpX+
clpX-
DksA remaining
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
clpX-
clpX+
0 5 10 15 20
time (min)

Figure 3.2. Western blots of trapped proteins.

(a) The molecular weights of bands for Dps (18.5 kDa), Rsd (18.1 kDa) and DksA (17.3 kDa) correspond to full-length proteins (F). The molecular weight of the RseA band (13 kDa) corresponds to an N-terminal fragment (N). The LexA fragments have masses (9 and 13 kDa) expected for autocleavage fragments consisting of residues 1-84 (N) and 85-202 (C). No immunoreactivity was observed in samples trapped in a *clpX*$^-$ strain.

(b) ClpX-dependent degradation *in vivo*. Following dilution from a stationary phase culture, protein synthesis was inhibited with spectinomycin at an A_{600} of 0.1, and samples were removed at specific time points and assayed by western blotting with anti-Dps or anti-DksA antibodies as indicated.

trapping as a method for global substrate discovery and suggest that most other captured proteins will also prove to be authentic ClpXP substrates.

Many trapped substrates have C-terminal degradation signals.

ClpX recognizes the C-terminal residues of certain substrates, including Leu-Ala-Ala-COOH of the ssrA tag and Arg-Arg-Lys-Lys-Ala-Ile-COOH of MuA (Levchenko et al. 1997b; Flynn et al. 2001). Inspection of the C-termini of the proteins trapped in a ClpX-dependent fashion revealed that 45% had sequences similar to either the ssrA tag or MuA (C-motifs 1 and 2 in Fig. 3.3a). Four trapped proteins had Ala-Ala terminal dipeptides, which in the ssrA tag is largely responsible for ClpX recognition (Flynn et al. 2001). Other trapped proteins had non-polar C-terminal dipeptides and basic side chains in the region 3 to 6 residues before the C-terminus. Positively charged residues at these positions are important for ClpX recognition of MuA (Levchenko et al. 1997b).

ClpX binding to the C-terminal sequences from Crl, RibB, LldD, YdaM and YbaQ was tested by inhibition of ClpXP degradation of GFP-ssrA (Fig. 3.3b, inset). Synthetic peptides corresponding to the 11 C-terminal amino acids of each of these proteins inhibited degradation of GFP-ssrA (Fig. 3.3b). Controls confirmed the specificity of this inhibition; neither an ssrA-peptide variant with Asp-Asp-COOH (Gottesman et al. 1998) nor the C-terminal peptide of Dps, which is not similar to either the ssrA or MuA tags, affected degradation of GFP-ssrA. Hence, the C-terminal residues of a number of proteins captured in a ClpX-dependent fashion bind ClpX, as expected for sequences that function as recognition signals.

To test directly for functional recognition, we fused the 10 C-terminal residues of Crl, Gcp and YbaQ to a stable reporter protein—Arc repressor—and assayed ClpXP degradation *in vitro*. Each fusion protein but not the parent Arc protein was rapidly degraded (Fig. 3.3c). Thus, these C-terminal sequences function as ClpXP-degradation signals. By extension, we suggest that most if not

a.

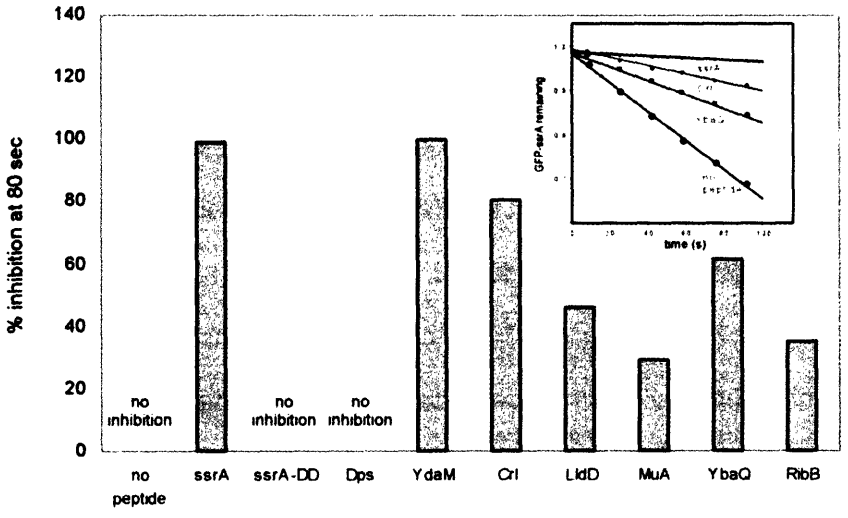
C-motif 1

ssrA	LAA-COOH
YdaM*	LAA
GlpD	LAS
Crl*†	LTA
PepB	LTA
Tpx	LKA
RplU	ISA
DksA	MAG
IscR	LRA
RplJ	EAA
LldD*	NAA
Gcp†	PAA
KatE	IPA
AldA	LQS
NrdH	ASA
CysA	QSA
AcnB	TAV
GlyA	VYA
CheW	VA
Fnr	VA
YcbW	AV
ExbB	AG

C-motif 2

MuA	RRKKAI-COOH
YbaQ*†	RAKKVA
RibB*	H RKAS
PncB	HIKKAS
Rsd	RVKHPA
PaaA	HARKVA

b.



c.

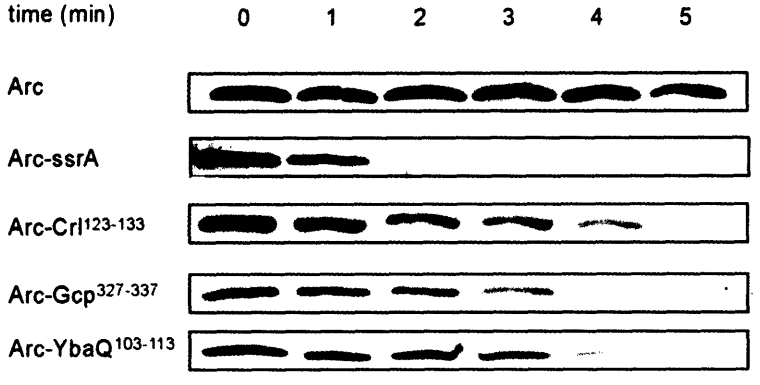


Figure 3.3. C-terminal recognition signals in trapped proteins.

(a) Sequence similarities of trapped proteins with the ssrA tag (C-motif 1) and MuA (C-motif 2). Dissimilar amino acids are shadowed in gray. * — proteins whose corresponding C-terminal peptides inhibit ClpXP degradation of GFP-ssrA. † — proteins whose C-terminal peptides target Arc-fusion proteins for ClpXP degradation.

(b) ClpXP degradation of GFP-ssrA in the presence of C-terminal peptides. Bars indicate percent inhibition after 80 sec of degradation from experiments like those shown in inset. Peptide sequences were ssrA (CAANDENYALAA), ssrA-DD (CAANDENYALDD), Dps (CFLWFIESNIE), YdaM (CKNDGRNRVLAA), Crl (CDFRDEPVKLTA), LldD (CALAPMAKGNAA), MuA (CILEQNRRKKAI), YbaQ (CARREERAKKVA), and RibB (CAYRQAHERKAS).

(c) ClpXP degradation of Arc fusion proteins with the ssrA tag or C-terminal residues of Crl (FRDEPVKLTA), Gcp (RWPLAELPAA), and YbaQ (RREERAKKVA) assayed by SDS-PAGE.

all of the proteins listed in Fig. 3.3a have C-terminal peptide signals that make them substrates for ClpXP.

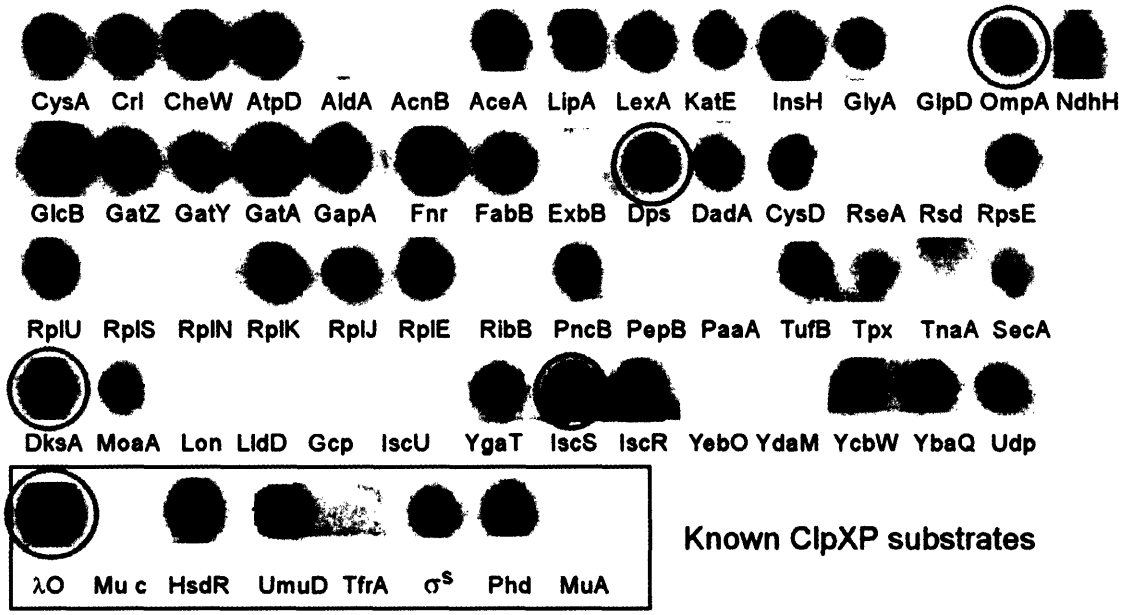
Peptide arrays identify N-terminal ClpX-binding signals.

To test for potential N-terminal ClpX-recognition signals, we prepared a peptide array with the N-terminal 11 residues of the ClpXP-trapped proteins and several previously identified ClpXP substrates attached covalently to a filter. This array was incubated with ClpX and ATP□S, washed, and peptide-associated ClpX was detected with anti-ClpX antibody (Fig. 3.4a). ClpX bound to the N-terminal peptides of about 60% of the proteins tested. The specificity of peptide binding was evident from inspection of the filter; ClpX-binding ranged from very strong to undetectable. Notably, ClpX bound strongly to the N-terminal peptide of λO, a protein whose N-terminal residues are known to be important for ClpXP degradation (Gonciarz-Swiatek et al. 1999). These results suggest that ClpX may recognize many trapped proteins through N-terminal signals.

Alignments reveal multiple classes of N-terminal recognition motifs.

Inspection of the N-terminal sequences bound by ClpX revealed several distinct motifs. For instance, λO, Dps, and sixteen other trapped proteins contained good matches to the consensus: polar-T/Φ-Φ-basic-Φ where Φ indicates a hydrophobic side chain (N-motif 1 in Fig. 3.4b; also see Table 3.1). As an example of an N-motif-1 protein, we studied Dps. Purified Dps was efficiently degraded in a reaction requiring ClpX, ClpP, and ATP (Fig. 3.5a; data not shown). In contrast, a truncated Dps variant missing most of N-motif 1 (Dps^{6-167}) was resistant to ClpXP degradation (Fig. 3.5a). Thus, the N-terminal residues of Dps are required for its degradation by ClpXP. These residues are absent in the Dps crystal structure (Grant et al. 1998), suggesting that they are unstructured and would therefore be accessible to ClpX. A deletion variant of λO missing N-motif 1 is also less susceptible to ClpXP degradation (Gonciarz-Swiatek et al. 1999), supporting a role for this sequence in ClpX recognition of λO.

a.


CysA Crl CheW AtpD AldA AcnB AceA LipA LexA KatE InsH GlyA GlpD OmpA NdhH
GlcB GatZ GatY GatA GapA Fnr FabB ExbB Dps DadA CysD RseA Rsd RpsE
RplU RplS RplN RplK RplJ RplE RibB PncB PepB PaaA TufB Tpx TnaA SecA
DksA MoaA Lon LldD Gcp IscU YgaT IscS IscR YebO YdaM YcbW YbaQ Udp
λO Mu c HsdR UmuD TfrA σS Phd MuA
Known ClpXP substrates

b.

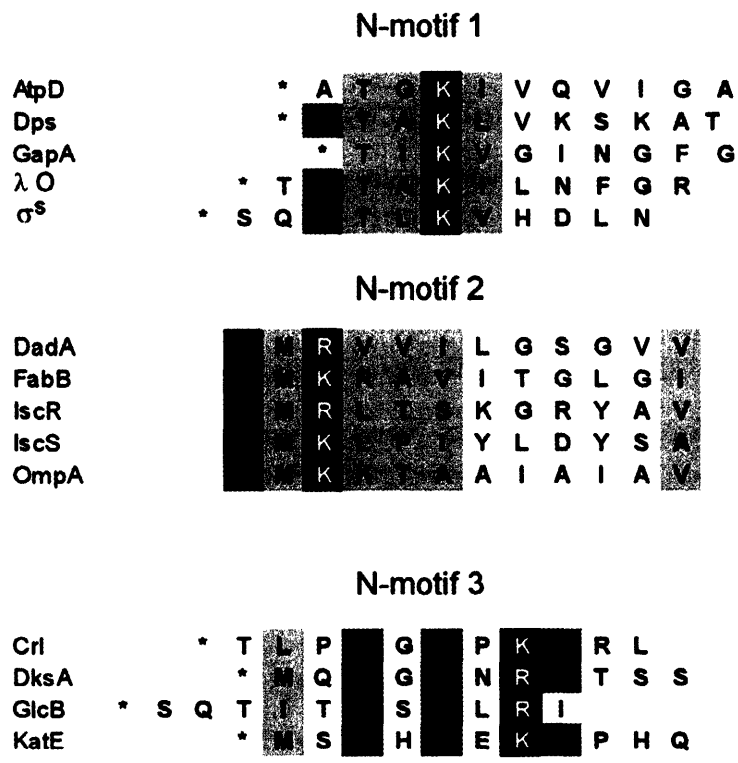

N-motif 1
AtpD * A K V Q V I G A
Dps * K V K S K A T
GapA * K G I N G F G
λO * T K L N F G R
σS * S Q K H D L N
N-motif 2
DadA R L G S G V
FabB K I T G L G
IscR R K G R Y A
IscS K Y L D Y S
OmpA K A I A I A
N-motif 3
Crl * T L P G P K R L
DksA * Q G N R T S S
GlcB * S Q T T S L R I
KatE * S H E K P H Q

Figure 3.4. N-terminal recognition signals.

(a) A filter with covalently bound peptides corresponding to the N-terminal 11 residues of trapped proteins and known ClpXP substrates was incubated with ClpX and bound protein was detected as in a western blot (see Experimental Procedures). Removal of the N-terminal Met was assumed for proteins with Ala, Ser, Thr, or Gly at position 2 and peptides corresponded to residues 2-12 of the unprocessed molecule (Ben-Bassat et al. 1987). Peptides shown to target fusion proteins for ClpXP degradation are circled.

(b) Many ClpX-binding sequences contain one of three motifs: N-motif 1: polar-T/Φ-Φ-basic-Φ; N-motif 2: NH_2-Met-basic-Φ-Φ-Φ-X_5-Φ; or N-motif 3: Φ-X-polar-X-polar-X-basic-polar. Additional members of each group are listed in Table 3.1. Asterisks correspond to the α-amino group.

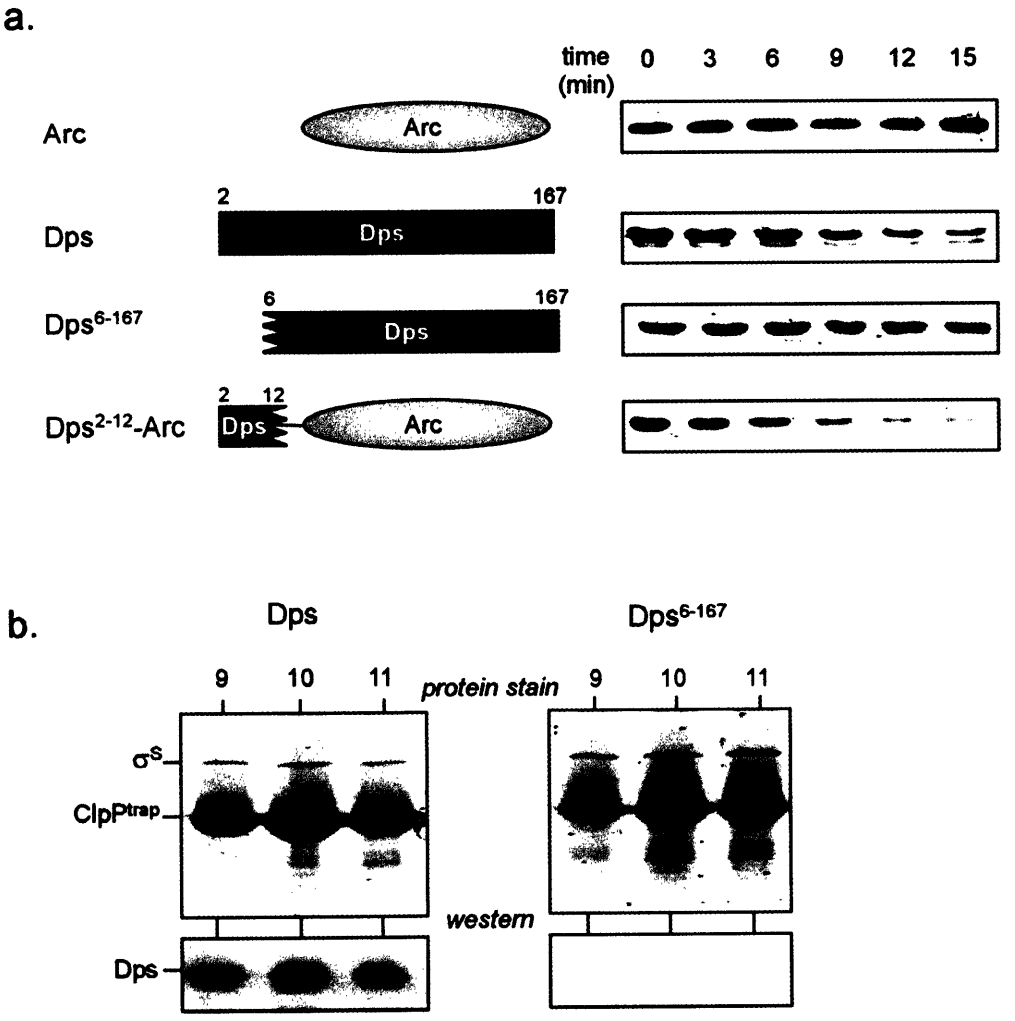


Figure 3.5. Dps has an N-terminal degradation signal.

(a) ClpXP-degradation of full-length Dps, full-length Arc, Dps^{6-167} or Dps^{2-12}-Arc assayed by SDS-PAGE.

(b) Purification of $ClpP^{trap}$ complexes formed in strains expressing Dps or Dps^{6-167}. $ClpP^{trap}$ was purified by Ni-NTA followed by gel filtration. The three peak $ClpP^{trap}$ fractions (9-11) are shown: (*upper panel*) stained with Sypro orange; (*lower panel*) probed with anti-Dps antibody. Note the presence of σ^S in the upper panel confirms that trapping occurred efficiently in both strains.

To establish that N-motif 1 is a functional ClpX-recognition signal *in vivo*, we co-expressed $ClpXP^{trap}$ with either Dps or Dps^{6-167}. As expected, full-length Dps co-purified with $ClpP^{trap}$ (Fig. 3.5b) but the truncated variant, Dps^{6-167} did not (Fig. 3.5b). These data demonstrate that N-motif 1 is essential for Dps-ClpX interactions in the cell.

To determine the sufficiency of the N-motif-1 sequence for ClpXP degradation, we constructed Arc fusion proteins containing the first 12 residues of Dps or λO. Following cellular removal of the N-terminal methionines, the purified proteins produced were Dps^{2-12}-Arc and λO^{2-12}-Arc. ClpXP degraded both fusion proteins *in vitro* at rates similar to those observed for full-length Dps and λO (Fig. 3.5a; Fig. 3.6a). Thus, the N-terminal regions of Dps and λO contain sequences that are both necessary and sufficient to target proteins for degradation by ClpXP.

Next, we mutated conserved residues in N-motif 1. Dps^{2-12}-Arc fusion proteins containing Asp substitutions for Thr3, Lys5, or Leu6 were degraded significantly less efficiently by ClpXP than the parental Dps-Arc fusion (Fig. 3.6a). These data establish that several of the conserved residues in N-motif 1 are important for its function as a ClpX-recognition signal.

DadA, IscS, OmpA, and nine additional proteins shared N-terminal sequences matching the pattern NH_2-Met-basic-Φ-Φ-Φ-X_5-Φ (N-motif 2 in Fig. 3.4b; Table 3.1). Adding either the $OmpA^{1-11}$ or $IscS^{1-11}$ sequences to the N-terminus of Arc converted it into a substrate for ClpXP degradation (Fig. 3.6b). Mutating Lys2 or Ile5 of the $IscS^{1-12}$ sequence to Asp abolished detectable degradation of the fusion protein, showing that these residues are essential for ClpX recognition of this sequence motif (Fig. 3.6b).

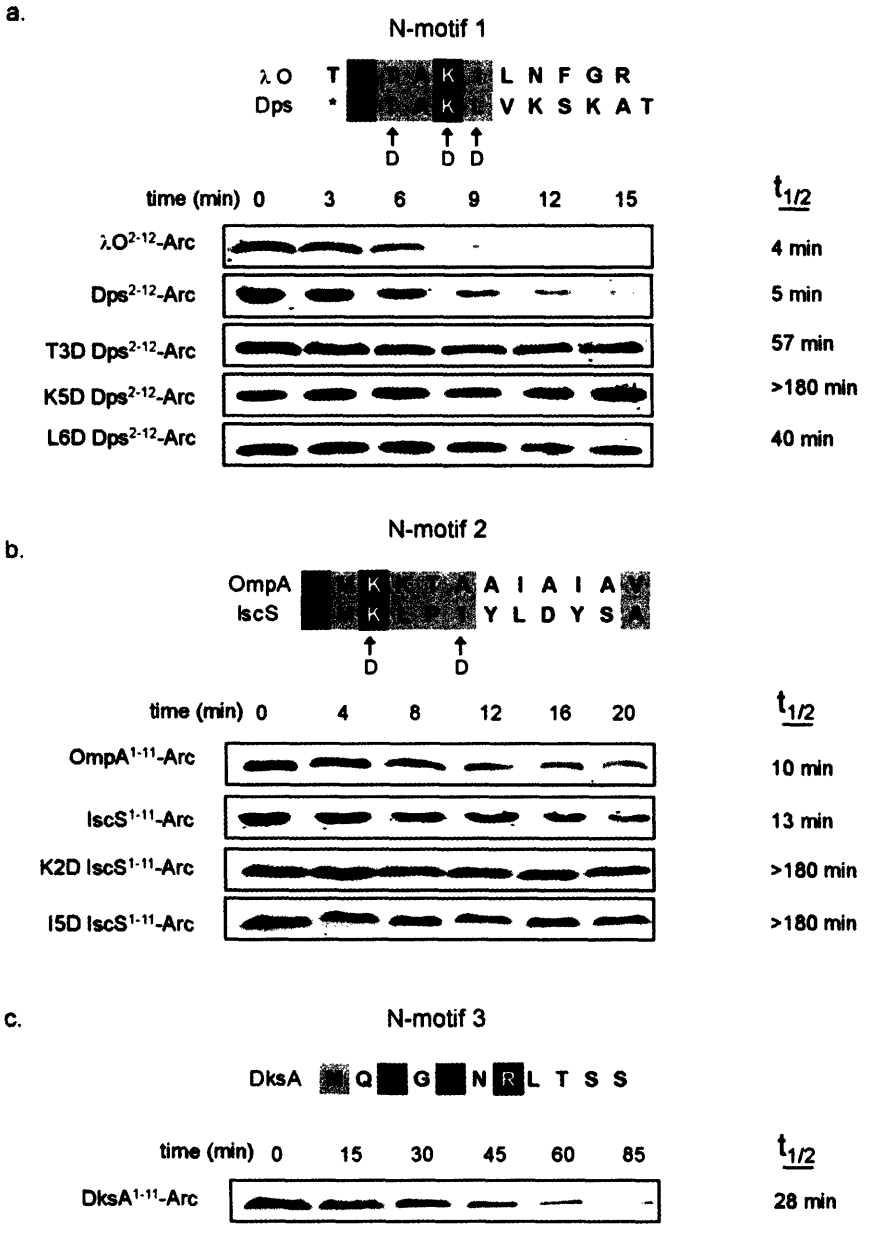


Figure 3.6. ClpXP degradation of Arc-fusion proteins with wild-type or mutant N-terminal recognition signals.

Degradation of each protein (5 μM) was assayed by SDS-PAGE and half-lives ($t_{1/2}$) were determined from plots of intensity versus time.

Ten other proteins, including Crl and DksA, contained N-terminal sequences that generally fit the consensus Φ-X-polar-X-polar-X-basic-polar (N-motif 3 in Fig. 3.4b; Table 3.1). When $DksA^{1-12}$, a representative sequence containing this motif, was fused to Arc, the resulting protein was degraded by ClpXP (Fig. 3.6c), although less rapidly than fusion proteins carrying N-motif-1 or N-motif-2 signals. Thus, representative sequences containing each of the three N-motifs were sufficient to confer susceptibility to degradation by ClpXP. These N-motifs represent new and distinct classes of ClpX-recognition signals.

Discussion

Substrate discovery through intracellular trapping.

Targeted protein degradation in bacteria is a dynamic process in which substrates of proteases like ClpXP change as cells respond to shifts in nutrients and to environmental stress. As a result, studying the full impact of degradation on the bacterial proteome requires methods for identifying protease substrates under a variety of environmental conditions. Here, we have described the use of an inactive, epitope-tagged variant of the ClpP protease as an intracellular trap for ClpXP substrates. Following capture and affinity purification, tandem-mass spectrometry identified more than 50 *E. coli* proteins. Similar strategies could be applied to identify protein targets of ClpXP under different growth conditions in *E. coli* or in other bacteria. Similar methods should also work to identify substrates of the ClpAP, HslUV, and Lon proteases.

Several observations support the conclusion that most ClpXP^{trap}-captured proteins are authentic ClpXP substrates. First, their capture by ClpP^{trap} depended on the presence of ClpX. Second, two known ClpXP substrates—σ^S and GFP-ssrA—were captured. Third, five newly identified trapped proteins (DksA, Dps, Fnr and two fragments of LexA) were subsequently shown to be substrates for ClpXP degradation. Fourth, the majority of ClpXP^{trap}-captured proteins displayed C-terminal and/or N-terminal peptide sequences that bound to ClpX or were very similar to known recognition signals and seven of the peptides identified in this manner were shown to target fusion proteins for ClpXP degradation. This collection of proteins captured by ClpXP^{trap} represents a large increase in the number of known ClpXP substrates.

For a few ClpXP^{trap}-associated proteins the relevance to ClpXP-mediated degradation was uncertain. For example, DnaK was also associated with ClpP^{trap} in the absence of ClpX. Because DnaK binds unfolded proteins (Pelham 1986), we assume that it binds denatured or unassembled ClpP^{trap} subunits. Hence, we have no evidence that DnaK is a ClpXP substrate. For OmpA, questions arose because the captured protein is normally located in another compartment, the outer

membrane. OmpA is highly expressed, however, and may saturate the SecA-mediated secretion pathway under some circumstances; ClpXP degradation of this cytoplasmic OmpA could play a role in protein-quality control. For RseA, we found that $ClpXP^{trap}$ captured an N-terminal fragment corresponding to its cytoplasmic domain whereas neither its C-terminal periplasmic domain nor the full-length protein, which spans the inner membrane, was trapped. Specific trapping of this N-terminal RseA domain supports a model proposed by Alba et al. (2002) in which ClpXP-mediated degradation of the N-terminal domain of RseA requires prior cleavage of RseA by inner-membrane proteases.

Seven proteins captured by $ClpXP^{trap}$ had masses ranging from 50 to 102 kDa even though structural calculations suggest the ClpP chamber can only accommodate globular proteins as large as 50 kDa (Wang et al. 1997; Ortega et al. 2000). How might these larger proteins be trapped? EM images of $ClpXP^{trap}$-substrate complexes reveal substrate density both within the ClpP chamber and at the axial ends of ClpXP particles (Ortega et al. 2000), suggesting that captured proteins can be associated with $ClpP^{trap}$ with only a portion of the substrate inside the chamber.

Molecular definition of ClpX-recognition motifs.

Identification of cellular proteins captured by $ClpXP^{trap}$ led to the discovery of five peptide motifs that target proteins for ClpXP degradation. Overall, nearly 90% of the proteins captured by $ClpXP^{trap}$ contain sequences that are attractive candidates for ClpX-recognition signals. Twenty-six of the captured proteins have C-terminal sequences that are plausible sites of ClpX interaction based on their similarities to known recognition signals, peptide-inhibition studies, and fusion protein analysis. These sequences fall into two classes; C-motif 1 is ssrA-like and C-motif 2 is more similar to the MuA-recognition sequence. In addition, forty of the captured proteins have N-terminal peptides that bound ClpX on a peptide array. Alignments of the N-terminal ClpX-binding sequences reveal three peptide motifs. Representative sequences from each of these motifs convert an attached protein into a

ClpXP substrate, demonstrating that these sequences are functional ClpX-recognition signals. Single point mutations in highly conserved motif residues also stabilize these fusion proteins, confirming the importance of these determinants for recognition. Thus, analyzing a large group of new ClpXP substrates has allowed us to define sequence rules governing substrate choice.

The ClpX-recognition motifs were clearly enriched in the trapped population of proteins compared to the entire proteome. For example, the percentage of trapped proteins terminating with the dipeptide Ala-Ala-COOH (the critical region of C-motif 1) was enriched seven-fold. N-motif 1 is the most defined of the three N-terminal recognition motifs. A strict consensus for this motif —T^1-X^2-K^3-[ILV]^4 located from one to four residues from the N-terminus—is present in the trapped protein population at a ten-fold higher frequency than in the proteome. Despite inherent uncertainties about whether these sequences will be accessible or functional in any specific protein, the identification of five classes of defined ClpX-recognition signals provides a useful foundation for the bioinformatic identification of other likely ClpX substrates.

In bacteria, many proteins are degraded by more than one protease. For example, ssrA-tagged proteins are degraded by ClpXP, ClpAP and FtsH, whereas SulA is degraded by HslUV (ClpYQ) and Lon (Gottesman et al. 1998; Herman et al. 1998; Wu et al. 1999). Some of the new ClpXP substrates identified here are also substrates for other proteases. For example, the C-terminal autocleavage fragment of LexA is degraded by ClpXP (Neher et al, 2003a) but is also a substrate for the Lon protease (Little 1983). Likewise, both ClpXP and other proteases appear to contribute to the degradation of Dps and DksA. Finally, a preliminary analysis of the proteins captured by $ClpP^{trap}$ in a strain expressing ClpA but not ClpX indicates that ClpAP recognizes about 10 proteins that are also recognized by ClpXP.

How most shared substrates are recognized by multiple proteases is not presently known. In the case of ssrA-tagged proteins, it has been established that the same 11-residue peptide targets them to ClpXP and to ClpAP, but it is also known that these proteases recognize a different set of amino acid residues within this peptide (Gottesman et al. 1998; Flynn et al. 2001). We believe that it

is also likely that ClpXP and ClpAP will recognize non-identical recognition signals in other shared substrates. Current evidence supports the idea that the precise peptide motifs that target proteins for degradation by ClpXP and ClpAP are different. For example, the shared substrates identified include proteins with N-motif 1 and N-motif 2 (see Table 3.1) but most N-motif 1 or N-motif 2 proteins are not common substrates. Furthermore, *in vitro* degradation experiments demonstrate that Dps, which is recognized by ClpXP via N-motif 1, is not degraded by ClpAP (unpublished data), indicating that this signal is not recognized by both proteases. Similarly, ClpA does not recognize C-motif 1 in the ssrA tag or C-motif 2 in MuA (Flynn et al. 2001; I. Levchenko and TAB, unpublished), and thus it is unlikely to directly recognize similar sequence motifs in other proteins.

In some instances, a ClpX-recognition signal normally located at a protein terminus can also function at some internal positions (Hoskins et al. 2002). However, analysis of previously characterized substrates and those described here suggests that ClpX-recognition signals are most commonly found near either the N-terminus or C-terminus of a protein. This localization is probably explained by the observation that these regions are frequently accessible in native proteins. Moreover, the free α-amino and α-carboxyl groups at the protein termini provide additional unique recognition determinants.

For LexA repressor, there is good evidence that an efficient ClpX-binding sequence is not recognized in the context of the full-length native protein. LexA contains an N-motif-2 sequence, which bound ClpX on the peptide array, but full-length LexA was neither captured by $ClpXP^{trap}$ nor degraded by ClpXP (Neher et al. 2003a). Inspection of the LexA crystal structure shows that portions of its N-terminal motif are buried in the native protein (Luo et al. 2001). In fact, for LexA and for RseA, accessible ClpXP recognition signals appear only to be produced following initial cleavage by other proteases. Recognition of cryptic peptide signals that are exposed as a result of polypeptide cleavage or protein denaturation probably represents a general strategy used by ClpX to interact with some substrates. This may explain why some captured proteins lacked recognizable N-terminal or C-terminal ClpX-binding motifs (see Table 3.1).

About one quarter of the captured proteins contain potential ClpX-recognition signals at both the N-terminus and C-terminus. In these cases, both signals might be utilized for ClpXP degradation or one or the other might be more accessible in the native protein or in protein complexes and therefore be used to a greater extent. In fact, precedence for multiple signals contributing to a protein's recognition by ClpX is evident from deletion analysis of the λO protein, which reveals that information located near both its N- and C-termini contributes to the efficiency of its degradation (Gonciarz-Swiatek et al. 1999). Even though some $ClpXP^{trap}$-captured proteins appear to have recognition signals at both the N-terminus and C-terminus, it seems unlikely that two ClpXP enzymes would ever degrade a single substrate from both ends, because the recognition signals bind rather weakly to ClpX hexamers and thus the probability that two ClpXP enzymes would simultaneously engage one substrate molecule is very low.

This study has revealed the presence of five classes of ClpX-recognition signals. In addition, one protein whose N-terminal peptide bound ClpX did not contain a recognizable motif, suggesting that there may be additional classes of signals. Why are there so many different types of signals? One attractive model is that signal diversity allows differential regulation of protein degradation. For example, proteins that bind specifically to one of the recognition motifs could specifically repress ClpXP degradation of these proteins but not those bearing other signals. As some single proteins appear to carry distinct classes of recognition signals, possibilities for combinatoral control of protein turnover are also present. It is common for multiple regulatory proteins to work together to control gene expression, and similar strategies could also help to regulate the precise composition of the proteome by degradation.

Trapped proteins and roles for ClpXP-mediated degradation.

Many of the proteins captured by $ClpXP^{trap}$ are co-regulated in response to cellular stress and changes in environment. For example, our analysis suggests that ClpXP degrades a set of proteins that are active during stationary phase. Five trapped proteins (Rsd, Dps, KatE, FtsZ, and GlpD) were encoded by genes transcribed under control of the stationary-phase σ^S factor, two additional captured proteins (Crl and DksA) have been implicated in controlling the level of σ^S, and σ^S itself represented one of the major trapped proteins (see Hengge-Aronis 199b and references therein; also see Pratt and Silhavy 1998; Jishage and Ishihama 1999; Webb et al. 1999). ClpXP is known to regulate σ^S levels by degrading it during exponential phase but not during stationary phase (Schweder et al. 1996). Our experiments indicate that Dps and DksA are degraded by ClpXP as the cells recover from stationary phase and re-enter logarithmic growth. Hence, ClpXP appears to regulate the levels of other stationary-phase proteins by direct degradation as well as by degrading σ^S.

Many proteins trapped by ClpXP help cells cope with oxidative stress and shifts between aerobic and anaerobic growth. Nine of the trapped proteins—Fnr, AceA, AcnB, AldA, GlcB, GlpD, MoaA, Tpx, and LldD—are encoded by genes regulated by the anoxic transcriptional regulatory proteins Fnr and/or ArcA (see Lynch 1996 and references therein; also see Kim et al. 1999; Pellicer et al. 1999a; Pellicer et al. 1999b; Anderson et al. 2000). Some oxidative stress probably occurred during our trapping experiments, as aerobic metabolism reduces O_2 to reactive species. Six trapped proteins—Fnr, IscR, IscU, AcnB, MoaA, and LipA—contain Fe-S centers, which can serve as sensors of oxidative stress. For example, the Fe-S cluster of Fnr is oxidized during aerobic growth (Kiley and Beinert 1998), reducing Fnr activity and potentially enhancing its degradation by ClpXP. Based on these initial studies, ClpXP may degrade proteins whose Fe-S clusters have been damaged by oxidation as a general response to oxidative stress.

Six ribosomal proteins were captured by $ClpXP^{trap}$. Why should proteins—such as ribosomal proteins—that are generally long-lived, be ClpXP substrates? Ribosome populations are reduced

following a nutritional downshift (Davis et al. 1986) and ClpXP may degrade ribosomes when nutrients become limiting, releasing amino acids for new protein synthesis. It is possible that ribosome turnover had begun when cells were harvested for our trapping studies during late exponential growth. Alternatively, ClpXP may degrade unassembled ribosomal proteins or damaged subunits. In fact, we suspect that for a number of substrates, ClpXP may function to degrade only a fraction of the protein population depending upon damage, assembly state, or growth conditions.

The definition of ClpX-recognition signals and the apparent role of ClpXP degradation in a variety of stress responses provides a foundation for understanding strategies for regulating protein turnover. Because peptide signals are critical for degradation, the use of signal-binding partners that mask or enhance substrate recognition by ClpX is one useful regulatory strategy. Regulating the availability of cryptic recognition signals provides another way to control degradation in response to environmental change. For example, denaturation of proteins during heat shock or initial cleavage by other proteases could expose latent ClpX-recognition sequences. Identification of additional ClpXP substrates under a broad range of environmental conditions should permit further definition of the molecular mechanisms that contribute to the cellular control of targeted protein degradation.

Acknowledgements

Supported by NIH grant AI-16892 and HHMI. T.A.B. is an employee of HHMI. We thank Richard Burgess, John Little, Akira Ishihama, Diana Downs, and Carol Gross for antibodies, Steven Finkel for strains and advice, Patricia Kiley for communication of unpublished work, Igor Levchenko, Shari Specter, and Anthony Schwacha for technical expertise and discussions, and the Harvard Microchemistry Facility for mass spectrometry. We thank Deborah Siegele and members of the Baker and Sauer labs for help, advice, and comments on this manuscript.

Experimental Procedures

Solutions: TBS: 50 mM Tris-HCl (pH 7.5) and 125 mM NaCl. ClpX buffer: 50 mM HEPES-KOH (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 100 μM $ZnSO_4$ and 2 mM DTT. PBS: 150 mM NaCl, 20 mM, Na-phosphate (pH 7.3). TEV buffer: 50 mM Tris-HCl (pH 8.0), 0.5 mM EDTA and 1 mM DTT. PD buffer, S buffer, W20 buffer, Clp buffer, and W500 buffer are as described (Kim et al. 2000).

Proteins: Dps and Dps^{6-167} (Grant et al. 1998), GFP-ssrA (Yakhnin et al. 1998), ClpP (Kim et al. 2000), and Arc derivatives (Arc-st11 and the fusions) (Robinson and Sauer 1996) were purified as described. ClpX was purified using standard chromatographic methods; the protocol is available upon request.

Strains and plasmids: *E. coli* strains were grown in LB broth. The W3110 *clpP::cat ΔsmpB-1*, W3110 *clpP::cat clpA::kan ΔsmpB*, and W3110 *clpP::cat clpX::kan ΔsmpB* strains were derived from W3110 *ΔsmpB-1* (Karzai et al. 1999). From this strain, additional protease mutations (*clpA::kan, clpX::kan, and clpP::cat*) were introduced by P1 transduction. To generate the MC4100 *clpX::kan clpP::cat clpA::kan* strain, the *clpP::cat* allele was transduced into SG22178.

A plasmid expressing ClpP without the pro-peptide sequence (Δ1-13) was constructed by PCR amplification of the *clpP* gene, cleavage with SphI and BglII, and cloning into the *SphI-BglII* fragment of QE-70. The active-site S97A mutation was introduced using Quickchange (Qiagen) and appropriate primers to generate pYK162. The Myc_3-TEV-His_6 sequence was introduced on an oligonucleotide cassette between the BglII and HindIII sites of pYK162 to produce pJF105. The C-terminal appended tag is: DSILTHRNRS HHHHHHGGEN LYFQGAYTSG EQKLISEEDL NGEQKLISEE DLNGEQKLIS EEDLN. Strains used for trapping were: JF148 (MC4100 *clpX::kan clpP::cat clpA::kan/pJF105*), JF176 (W3110 *clpP::cat ΔsmpB-1/pJF105*), JF172 (W3110 *clpP::cat clpX::kan ΔsmpB-1/pJF105*) and JF162 (W3110 *clpP::cat clpA::kan ΔsmpB-1/pJF105*).

A plasmid expressing Dps^{6-167} was constructed by PCR amplification from strain SK101 (Martinez and Kolter 1997), cleavage with NdeI and BamHI, and cloning into the *NdeI-BamHI* fragment of pET3a (Novagen). A plasmid expressing *arc-st11* in pET-11a was constructed by PCR amplification of pET-28b-Arc-ssrA (Burton et al. 2001) and ligation into the *NheI-BamHI* fragment of pET-11a (Novagen) to form pET-11a-Arc-st11. The first 12 residues of Dps and λO and the first 11 residues of IscS, OmpA and DksA were fused to Arc-st11 by using oligonucleotide cassettes. The mature N-terminal sequences of the fusion proteins are: Dps^{2-12}-Arc: *STAKLVKSK AS***MGK**; λO^{2-12}-Arc: *TNTAKILNF GRAS***MGK;** $IscS^{1-11}$-Arc: *MKLPIYLDY S AS***MGK**; $OmpA^{1-11}$-Arc: *MKKTAIAIAV AS***MGK**; $DksA^{1-11}$-Arc: *MQEGQNRKTS S***MGK** (Dps, λO, IscS, OmpA, and DksA in italics, Arc in bold). The T3D, K5D and L6D Dps^{2-12}-Arc mutants and the K2D and I5D $IscS^{1-11}$-Arc mutants were constructed using oligonucleotide cassettes. The C-terminal 10 residues of Crl, Gcp and YbaQ were fused to Arc-st11 by PCR amplification of the Arc-st11 gene with primers containing the C-terminal sequence of each respective protein and ligation into the *NheI-BamHI* fragment of pET11a. The sequence of the C-terminal region of the resulting fusion proteins are: Arc-$YbaQ^{103-113}$: *QHD***RREERA KKVA**; Arc-$Crl^{123-133}$: *QHD***FRDEPV KLTA**; Arc-$Gcp^{327-337}$: *QHD***RWPLAE LPAA**. All constructs were confirmed by DNA sequencing.

A plasmid expressing Dps under control of the arabinose promoter (pJF119) was constructed by removal of the *dps* and *araC* genes from pBAD18-dps (Martinez and Kolter 1997) and cloning into the *AvaI-HindIII* fragment of pSU38. Dps^{6-167}-pSU38 was constructed by PCR amplification of dps^{6-167} from the dps^{6-167}-pET3a plasmid and ligation into the *EcoRI/XbaI* fragment of pBAD18. The dps^{6-167} and *araC* genes were cut from the resulting plasmid and cloned into the *AvaI-HindIII* fragment of pSU38 to form pJF121. pJF119 and pJF121 were then transformed into JF176.

Protein trapping in vivo: Strains JF148, JF162, JF172 and JF176 were grown in 4 L of LB/amp at 30°C to an A_{600} of 0.4, induced with 0.5 mM IPTG and grown for 2.5 additional hrs. Cells were harvested by centrifugation and resuspended in 3 ml S buffer per gram of cells. Following lysis by

French press, the lysate was centrifuged for 30 min at 25,000 x g, and the supernatant was added to 2.5 ml nickel-NTA resin (Qiagen) equilibrated in S buffer. After mixing for 2 hrs at 4°C, the resin was packed into a column, washed with 200 ml S buffer, 100 ml W20 buffer, and eluted with 5 ml W500 buffer. The Myc antibody affinity resin was generated by cross-linking 9E10 antibody to protein G agarose (Invitrogen) as described (Harlow 1988). The elutant from the nickel-NTA column was mixed with 1.5 ml of this resin equilibrated in PBS. After mixing for 2 hrs at 4°C, the beads were packed into a column and washed with 60 ml PBS, followed by 60 ml PBST (PBS + 0.1% Tween 20), and finally by 20 ml TEV reaction buffer. The slurry was then mixed with 1 ml TEV reaction buffer and 400 units of TEV protease (Gibco), and agitated at room temperature for 30 min. The released protein was collected and stored at –20°C.

Trapping of Dps and Dps$^{6-167}$ in vivo: Dps or Dps$^{6-167}$ was co-expressed with ClpP^{trap} under the same conditions as above, by the addition of 0.2% L-arabinose at the same time as the IPTG. ClpP^{trap}-complexes were purified on a Ni-NTA column as above followed by filtration chromatography on a Superdex 200 PC 3.2/30 column run in Clp buffer.

2D gels: Samples for 2D gel analysis were exchanged into 8 M urea and 2% CHAPS and loaded on a 7 cm Immobiline DryStrip (pH 3-10L) for focusing on a IPGphor system (Pharmacia), followed by 12.5% SDS-PAGE (Bjellqvist et al. 1993). Spots were visualized using Sypro Ruby protein stain (Molecular Probes) on a Fluorimager 595 (Molecular Dynamics).

Mass spectrometry: Samples for MS/MS analysis were separated by 12.5% SDS-PAGE. Gel slices (approximately 0.5-1.0 cm) were excised, digested with trypsin, and analyzed by microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry using a Finnigan LCQ DECA quadropole ion trap mass spectrometer (Harvard Microchemistry Facility). Control analyses performed on samples purified from the *clpX*$^-$*clpA*$^-$ strain yielded peptides from: ClpP, TEV protease

and keratin, as well as 4 peptides of Dps. The presence of this small number of Dps peptides was probably an artifact due to purification of Dps in the laboratory during sample preparation; western analysis failed to detect any Dps in this sample (see Fig. 3.2a).

Degradation in vivo: Cultures of W3110 or W3110 *clpX::kan* cells were grown overnight in LB broth at 37 °C ($A_{600} \approx 3$), diluted 1:100 in fresh LB broth, and allowed to grow for 50 minutes at 37 °C ($A_{600} \approx 0.1$). At this point, 150 μg/ml of spectinomycin was added. Samples were removed at specific times and analyzed by SDS-PAGE, followed by western blotting (see below).

Western blots: Western blots were performed following the guidelines of Amersham for use with the ECF substrate (Amersham) using the following primary antibodies: anti-Dps (from Richard Bugess, University of Wisconsin, Madison), anti-LexA (from John Little, University of Arizona), anti-Rsd (from Akira Ishihama, National Institute of Genetics), anti-DksA (from Diana Downs, University of Wisconsin, Madison), or anti-N-domain RseA and anti-C-domain RseA (from Carol Gross, UCSF).

Degradation in vitro: $ClpX_6$ (0.3 μM), $ClpP_{14}$ (0.8 μM), ATP (4 mM), and an ATP regeneration system (50 μg/ml creatine kinase and 2.5 mM creatine phosphate) were mixed in PD buffer and incubated for two minutes at 30°C. For all degradation experiments 5 μM of protein added and samples were removed at specific times and analyzed by SDS-PAGE. For peptide-inhibition experiments, GFP-ssrA (1 μM) was added with peptide (50 μM) and degradation was monitored by fluorescence as described (Flynn et al. 2001).

Peptide arrays: A cellulose filter containing peptides corresponding to the 11 N-terminal residues of all the trapped proteins (except GroEL, FtsZ, ClpX and DnaK) and known ClpXP substrates was prepared by the MIT Biopolymers facility using an Abimed instrument. Each peptide contained two

additional C-terminal β-alanines, and was covalently attached to the filter by a polyethylene glycol linker. The filter was soaked in ethanol, washed three times for 5 min in TBST (TBS + 0.1% Tween 20), blocked overnight in TBST plus 10% milk, and then washed twice with TBST and twice in ClpX buffer for 5 min. $ClpX_6$ (0.8 μM) and ATPγS (4 mM) (Roche) were incubated at 30°C in 5 ml ClpX buffer for two min and added together with 0.1% milk to the filter for 6 hrs at 4°C. The filter was washed three times with ClpX buffer and ATPγS (0.5 mM) and incubated with anti-ClpX antibody in 5 ml ClpX buffer and ATPγS (1 mM) for 30 min. Next, the filter was washed three times as above, incubated with goat anti-rabbit IgG HRP-conjugated antibody (Amersham) and ATPγS (1 mM) for 20 minutes. After three final washes, the filter was incubated with ECL substrate (NEN), and visualized on film.

CHAPTER FOUR:

Modulating substrate choice: The SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation

---

This chapter is in press as Flynn, J.M., Levchenko, I., Sauer, R.T., and Baker, T.A. Genes and Development **18** (2004). Igor Levchenko purified cloned and purified σ^E, $RseA^{1-108}DD$, and provided advice on many experimental details. I. Levchenko, R.T. Sauer, and T.A. Baker were actively involved in preparing the manuscript.

Abstract

Adaptor proteins help proteases modulate substrate choice, ensuring that appropriate proteins are degraded at the proper time and place. SspB is an adaptor which delivers ssrA-tagged proteins to the AAA+ protease ClpXP for degradation. To identify new SspB-regulated substrates, we examined proteins captured by $ClpXP^{trap}$ in *sspB*$^+$ but not *sspB*$^-$ strains. $RseA^{1-108}$, a fragment of a trans-membrane protein that regulates the extracytoplasmic-stress response, fit this criterion. In response to stress, RseA is cleaved on each side of the membrane and is released as a cytoplasmic fragment that remains bound in an inhibitory complex with the σ^E transcription factor. Trapping experiments together with biochemical studies show that ClpXP functions in concert with SspB to efficiently recognize and degrade $RseA^{1-108}$ and thereby release σ^E. Genetic studies confirm that ClpX and SspB participate in induction of the σ^E regulon *in vivo*, acting at the final step of an activating proteolytic cascade. Surprisingly, the SspB-recognition sequence in $RseA^{1-108}$ is unrelated to its binding sequence in the ssrA tag. Thus, these experiments elucidate the final steps in induction of the extracytoplasmic-stress response and reveal that SspB delivers a broader spectrum of substrates to ClpXP than has been recognized.

Introduction

The AAA+ protease ClpXP performs a diverse array of cellular tasks, including degrading incomplete polypeptides, adjusting the activity of metabolic enzymes, and altering the levels of regulatory proteins in response to stress (Gottesman et al. 1998; Wang et al. 1999; Maurizi and Rasulova 2002; Flynn et al. 2003; Gottesman 2003). As a result, many substrates compete for degradation by a relatively small number of ClpXP protease molecules (Ortega et al. 2004). The priority of substrate recognition and degradation can also be controlled by adaptor proteins, which enhance or inhibit interactions between specific substrates and ClpXP or other AAA+ proteases (Dougan et al. 2002a). How widely adaptor proteins are used to control substrate choice is not currently understood.

In the ClpXP protease, ClpX—a hexameric-ring ATPase—binds native substrate proteins, denatures these molecules, and translocates the unfolded polypeptides into an internal degradation chamber of the ClpP peptidase (Maurizi et al. 1990; Wojtkowiak et al. 1993; Maurizi et al. 1994; Wang et al. 1997; Weber-Ban et al. 1999; Kim et al. 2000; Kim and Kim 2003). ClpX binds to short unstructured peptides called recognition signals or degradation tags, usually located near the N- or C-terminus of substrates (Levchenko et al. 1997; Gottesman et al. 1998; Gonciarz-Swiatek et al. 1999; Flynn et al. 2003). The ssrA degradation tag is a well-characterized 11-residue peptide (*AANDENYALAA*), which is added co-translationally to nascent polypeptides when ribosomes stall (Keiler et al. 1996). SsrA-tagging frees these distressed ribosomes for new rounds of translation and targets the incomplete polypeptides for degradation by ClpXP and other proteases (Gottesman et al. 1998; Withey and Friedman 2003).

The SspB adaptor was originally identified by its ability to enhance ClpXP degradation of ssrA-tagged proteins (Levchenko et al. 2000) and is one of the best-characterized proteins that functions in substrate delivery (Wah et al. 2002; Dougan et al. 2003; Levchenko et al. 2003; Song and Eck 2003; Wah et al. 2003; Bolon et al. 2004). SspB enhances recognition of

ssrA-tagged proteins by mediating the assembly of ternary complexes in which the substrate, adaptor, and protease are tethered by the following three sets of protein-peptide interactions: (1) the AAA+ domain of ClpX binds to the C-terminal *LAA* sequence of the ssrA tag; (2) the substrate-binding domain of SspB interacts with a sequence spanning the N-terminal seven residues of the ssrA tag; and (3) a short peptide sequence at the end of a flexible SspB tail binds directly to the N-terminal domain of ClpX (Levchenko et al. 2000; Flynn et al. 2001; Levchenko et al. 2003; Wah et al. 2003; Bolon et al. 2004). Whether SspB delivers any substrates without ssrA tags for ClpXP degradation has not been addressed.

Here, we show that SspB directs ClpXP recognition of *Escherichia coli* proteins, which are not ssrA-tagged. One of these substrates, RseA, functions as a master regulator of the extracytoplasmic-stress response by inhibiting the transcription factor (σ^E) that activates expression of stress genes (De Las Penas et al. 1997b; Missiakas et al. 1997; Dartigalongue et al. 2001; Rezuchova et al. 2003). RseA is a trans-membrane protein with an N-terminal cytoplasmic domain, which normally binds to and inhibits σ^E (De Las Penas et al. 1997b; Missiakas et al. 1997). In response to the stress-induced accumulation of unfolded or unassembled outer-membrane proteins in the periplasm, RseA is processed via multiple cleavage events in a sequential cascade. DegS protease initially cleaves RseA within its periplasmic domain, activating a second cleavage on the cytoplasmic side of the membrane by YaeL protease (Alba et al. 2001; Alba et al. 2002; Kanehara et al. 2002). These cleavage events release the cytoplasmic domain of RseA from the membrane, but this inhibitory domain remains bound to σ^E and thus additional steps are required before σ^E can activate gene expression (Missiakas et al. 1997; Campbell et al. 2003).

Our experiments demonstrate that ClpXP and SspB play a role in the final step of the proteolytic cascade that activates σ^E. Cleavage of RseA on the cytoplasmic side of the membrane generates a fragment ending in a ClpX-recognition signal, similar to the *LAA* sequence at the end of the ssrA tag. By binding simultaneously to this $RseA^{1-108}$ fragment and

ClpX, SspB brings the σ^{E}•$RseA^{1-108}$ complex and the ClpXP protease together. The RseA fragment is, however, the only component of this complex that is degraded. Surprisingly, the peptide sequences bound by SspB in $RseA^{1-108}$ and the ssrA tag are not similar, suggesting the SspB has different modes of protein recognition. These results establish that the SspB adaptor recognizes and delivers different classes of cellular proteins for degradation by ClpXP.

Results

SspB influences recognition of a set of ClpXP substrates in vivo.

To investigate whether SspB controls ClpXP degradation of proteins without ssrA tags, we compared intracellular substrates captured in an inactive variant of ClpP ($ClpP^{trap}$) in the presence and absence of SspB (Flynn et al. 2003). Trapping strains were *smpB*$^-$, which inactivates ssrA tagging (Karzai et al. 2000), and *clpA*$^-$, which removes another ATPase capable of choosing substrates for ClpP. These mutations eliminate trapping of ssrA-tagged and ClpAP substrates. Following capture in *sspB*$^+$ or *sspB*$^-$ strains, ClpXP substrates were visualized by staining following 2-D gel electrophoresis (Fig. 4.1). This experiment revealed that the majority of cellular substrates do not require SspB to interact with ClpXP. However, a handful of proteins were clearly more abundant in $ClpP^{trap}$ when SspB was present. This differential trapping indicates that SspB influences the recognition of a subset of ClpXP substrates. Interestingly, a few proteins were more efficiently trapped when SspB was absent, suggesting that SspB may also inhibit ClpXP degradation of certain proteins.

One SspB-dependent substrate is an N-terminal fragment of RseA.

Tandem-mass spectrometry identified one of the most prominent SspB-dependent ClpXP trapped proteins as an N-terminal fragment of RseA. Tryptic digestion of the RseA spot followed by mass spectrometry identified peptides covering the N-terminal 108 amino acids of RseA (Fig. 4.2), including a peptide with a molecular weight corresponding to residues 94 to 108: *VRPWAAQLTQMGVAA*^{108}. The fact that this peptide did not terminate with lysine or arginine (as expected for an internal tryptic fragment) indicated that alanine was the natural C-terminus of the trapped protein. Thus, this analysis demonstrates that the trapped RseA fragment ($RseA^{1-108}$) terminates with the sequence *VAA-COOH* (Fig. 4.2). This C-terminal sequence is a member of the well-characterized C-motif 1 class of ClpX-recognition signals (Flynn et al. 2003), and thus it makes sense that it would target the RseA fragment to ClpXP.

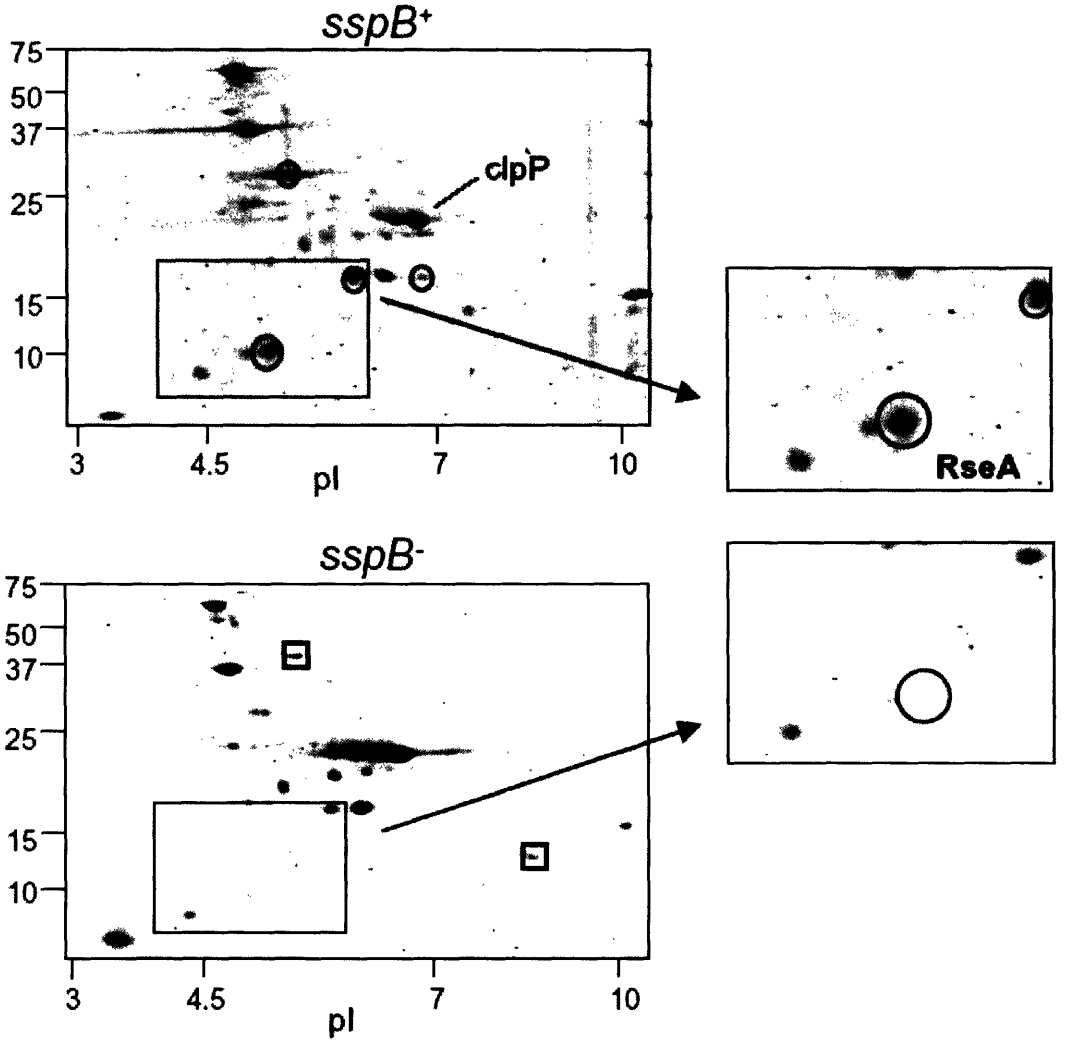


Figure 4.1. Proteins captured by ClpXP^{trap} with and without SspB.

2D-gel analysis of proteins captured by ClpXP^{trap} in *E. coli* strains JF162 (*sspB*$^{+}$*clpA*$^{-}$; top panel) and JF259 (*sspB*$^{-}$ *clpA*$^{-}$; bottom panel). Representative proteins trapped preferentially in the *sspB*$^{+}$ strain are circled whereas proteins trapped preferentially in the *sspB*$^{-}$ strain are marked by squares.

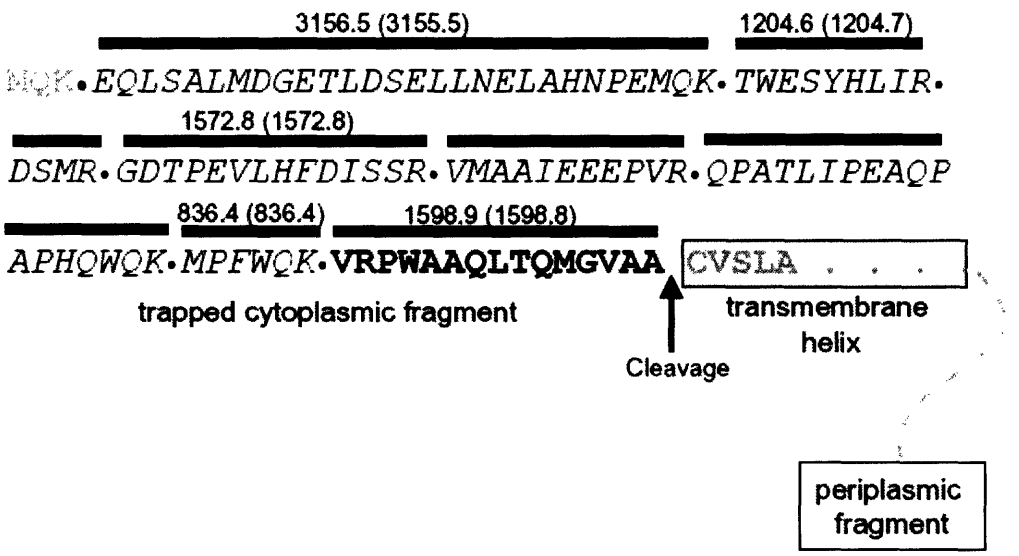


Figure 4.2. Sequence analysis of the ClpXP^{trap}-captured RseA fragment.

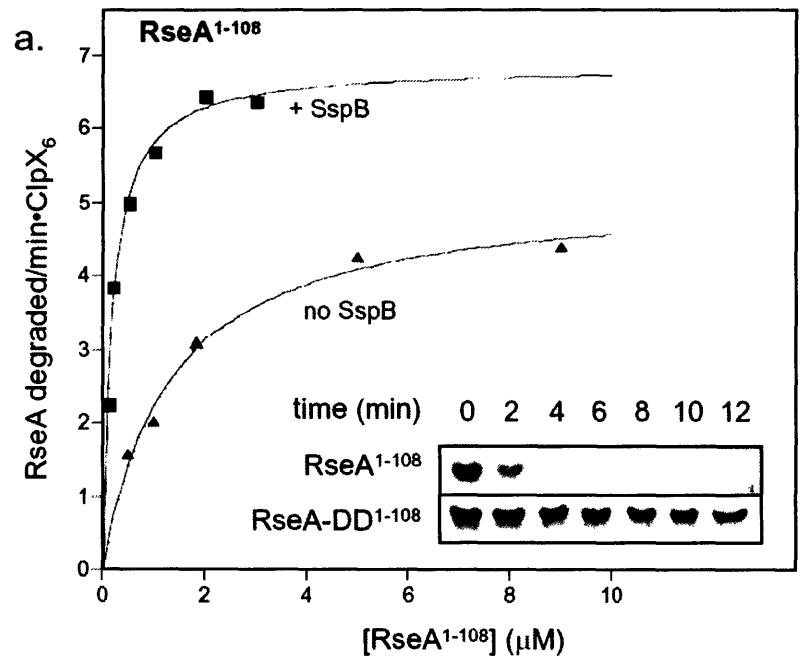
Tryptic fragments of the RseA fragment were analyzed by mass spectrometry. Identified peptides are marked with bold lines above the corresponding sequences; sequences identified by tandem mass spectrometry are italicized and peptides identified by MALDI mass spectrometry have the experimental (expected) molecular weights listed. The peptide highlighted in bold was identified by MALDI mass spectrometry and is the C-terminal tryptic peptide of the trapped fragment.

Recognition of RseA by the cytoplasmic protease ClpXP must occur after YaeL cleavage releases the N-terminal fragment from the membrane (Alba et al. 2001; Kanehara et al. 2002). Indeed, based on the cleavage specificity of the homologous SP2 protease, Alba *et al.* (2002) proposed that YaeL might cleave RseA between A^{108} and C^{109}, to generate the N-terminal fragment that we trapped and characterized.

$RseA^{1-108}$ is a substrate for SspB and ClpXP in vitro.

A fragment corresponding to $RseA^{1-108}$ was cloned, over-expressed, and purified to investigate its susceptibility to ClpXP degradation *in vitro*. ClpXP degraded $RseA^{1-108}$ in a reaction that required ATP (Fig. 4.3a; data not shown). A mutant variant in which the C-terminal sequence was VDD^{108} (RseA-DD^{1-108}) was also purified and was found to be degraded 25-30 times more slowly than $RseA^{1-108}$ (Fig. 4.3a, inset). Thus, we conclude that the C-terminal sequence of $RseA^{1-108}$ is a critical signal that targets this protein for degradation by ClpXP.

As expected from the trapping results, SspB also stimulated degradation of $RseA^{1-108}$ by ClpXP *in vitro* (Fig. 4.3). SspB reduced the Michaelis constant (K_m) for ClpXP degradation of $RseA^{1-108}$ approximately seven-fold from 1.3 to ~0.2 μM and stimulated V_{max} by ~50% (Fig. 4.3a). Thus, SspB enhances productive interactions between $RseA^{1-108}$ and ClpX, in a manner analogous to its role in delivering ssrA-tagged proteins for ClpXP degradation (Levchenko et al. 2000). A truncated SspB variant lacking the tails that bind ClpX did not enhance ClpXP degradation of $RseA^{1-108}$, demonstrating that tethering interactions between SspB and ClpX are important for delivery of this substrate (data not shown).


a.
RseA1-108
+ SspB
no SspB
time (min) 0 2 4 6 8 10 12
RseA1-108
RseA-DD1-108
RseA degraded/min•ClpX6
[RseA1-108] (μM)

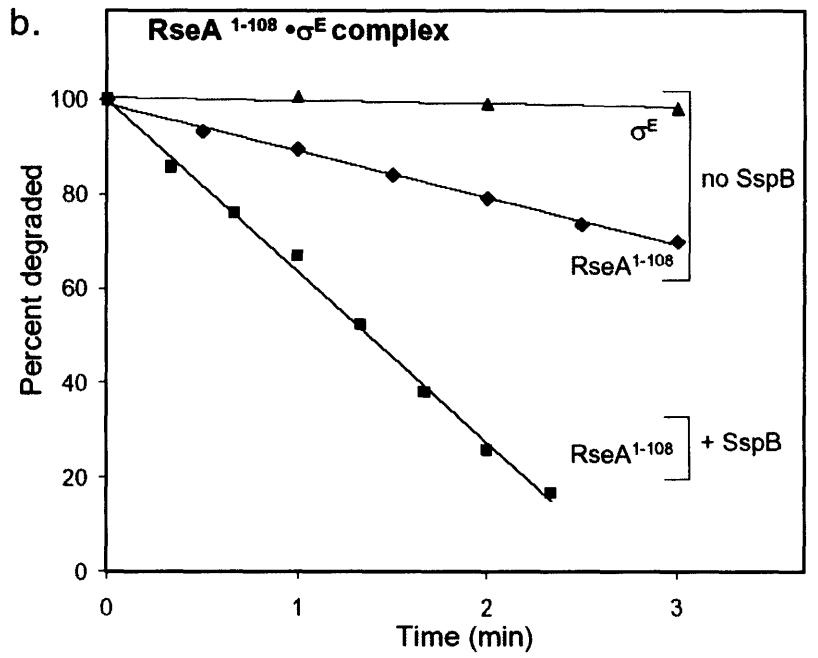

b.
RseA 1-108 •σE complex
σE
no SspB
RseA1-108
RseA1-108
+ SspB
Percent degraded
Time (min)

Figure 4.3. *ClpXP efficiently degrades purified RseA$^{1-108}$ in an SspB-stimulated manner.*

(a) Rates of ClpXP-mediated degradation of ^{35}S-labeled RseA$^{1-108}$ by $ClpX_6$ (50 nM) and $ClpP_{14}$ (150 nM) were determined at different substrate concentrations in the presence or absence of SspB (200 nM). Degradation was assayed by changes in TCA-soluble radioactivity, and rates were plotted against the substrate concentration. The solid lines are fits to the Michaelis-Menten equation in the absence (K_m = 1.3 μM, V_{max} = 5.2 min^{-1}) and or presence (K_m = 0.18 μM, V_{max} = 6.8 min^{-1}) of SspB.

Inset Degradation of RseA-DD$^{1-108}$ (2 μM) or RseA$^{1-108}$ (2 μM) by $ClpX_6$ (300 nM) and $ClpP_{14}$ (800 nM) was assayed by SDS-PAGE gel.

(b) ClpXP degradation of RseA$^{1-108}$ complexed with σ^E. ^{35}S-labeled RseA$^{1-108}$ (500 nM) was incubated with unlabeled σ^E (500 nM) for 5 min at 30 °C. Degradation by $ClpX_6$ (50 nM) and $ClpP_{14}$ (150 nM) was assayed by changes in TCA-soluble radioactivity in the presence (■) and absence (♦) of SspB (200 nM). ^{35}S-labeled σ^E was also incubated with unlabeled RseA$^{1-108}$ and ClpXP degradation was monitored in the same manner (▲). No detectable σ^E degradation by ClpXP was observed in the presence of SspB.

YaeL cleavage releases the N-terminal fragment of RseA from the membrane but does not disrupt its binding to σ^E (Alba et al. 2002; Kanehara et al. 2002). We asked, therefore, whether SspB could deliver the σ^E•RseA$^{1-108}$ complex to ClpXP for disassembly and degradation. As shown in Fig. 4.3b, ClpXP degraded RseA$^{1-108}$ bound to σ^E, and SspB stimulated this degradation. At the concentrations tested, the rate of degradation of free RseA$^{1-108}$ was similar to that of complexed RseA$^{1-108}$. This result indicates that binding of σ^E to RseA$^{1-108}$ does not inhibit degradation or provide any critical contacts that enhance recognition of RseA$^{1-108}$ by ClpX. Importantly, σ^E in the σ^E•RseA$^{1-108}$ complex was not degraded. In addition, as expected, SspB remained undegraded throughout the reaction (data not shown).

Based on this analysis we conclude that SspB can deliver the σ^E•RseA$^{1-108}$ complex to ClpXP, leading to the targeted degradation of RseA$^{1-108}$. These results are integrated into a model for σ^E activation shown in Figure 4.4. Following DegS and YaeL cleavage of RseA, SspB delivers the σ^E•RseA$^{1-108}$ complex to ClpX, which selectively denatures RseA$^{1-108}$ and translocates it into ClpP for degradation. This processing of the RseA fragment by ClpXP releases SspB and σ^E from the enzyme complex. As a consequence, σ^E is liberated to bind to core RNA polymerase and activate transcription.

SspB and ClpX enhance activation of the σ^E regulon in vivo.

Taken together, the results presented so far suggest that degradation mediated by ClpXP and SspB controls the intracellular levels of RseA$^{1-108}$ and should therefore influence σ^E activity. To test for roles for ClpX and SspB in the extracytoplasmic-stress response, we monitored induction of a σ^E-controlled *lacZ* reporter gene following induction of the stress response in *sspB*$^-$ or *clpX*$^-$ cells. Extracytoplasmic stress was induced using a plasmid-encoded fusion protein, ending with a *YYF* sequence, which is targeted to the periplasm and activates DegS degradation of RseA (Walsh et al. 2003). Following induction, σ^E-dependent β-galactosidase synthesis was delayed in both the *clpX*$^-$ and *sspB*$^-$ cells (Fig. 4.5). These data

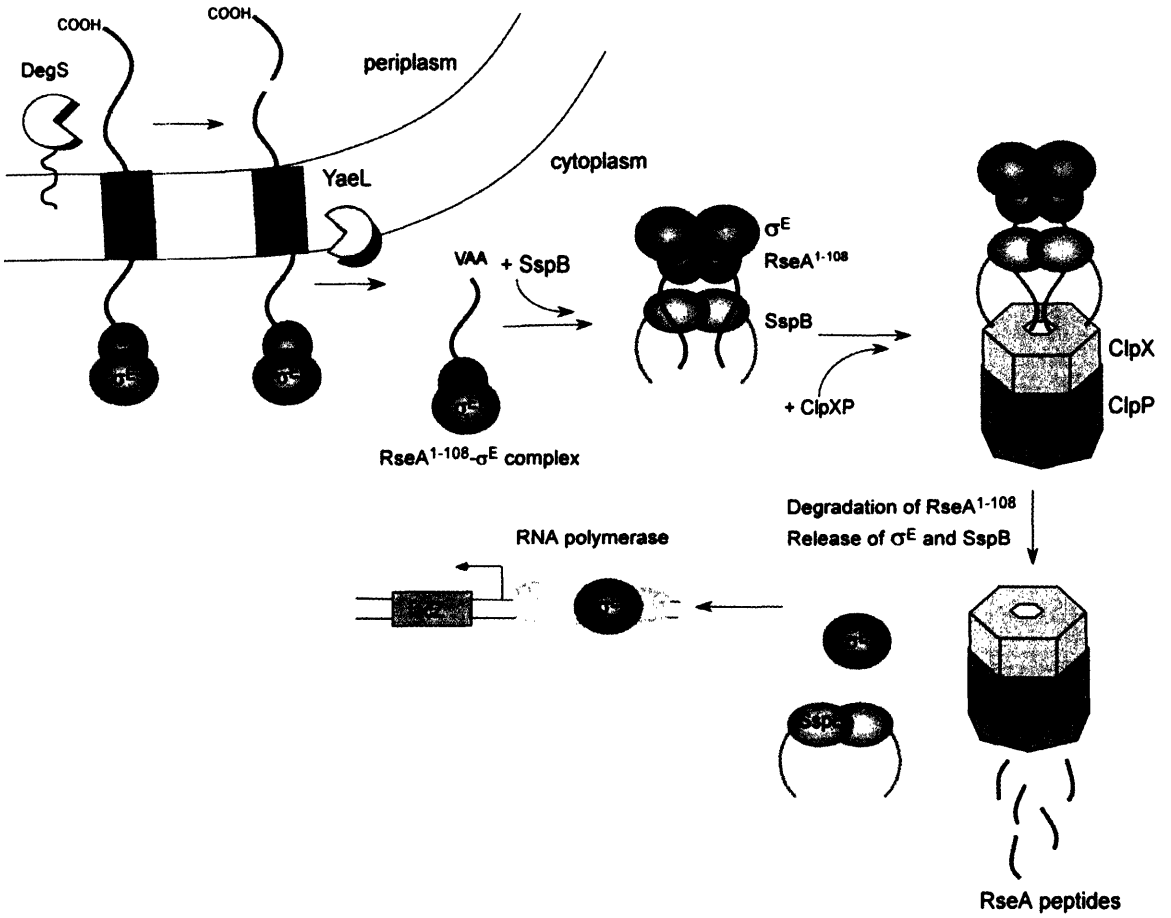


Figure 4.4. Activation of σ^E mediated by a cascade of RseA proteolysis.

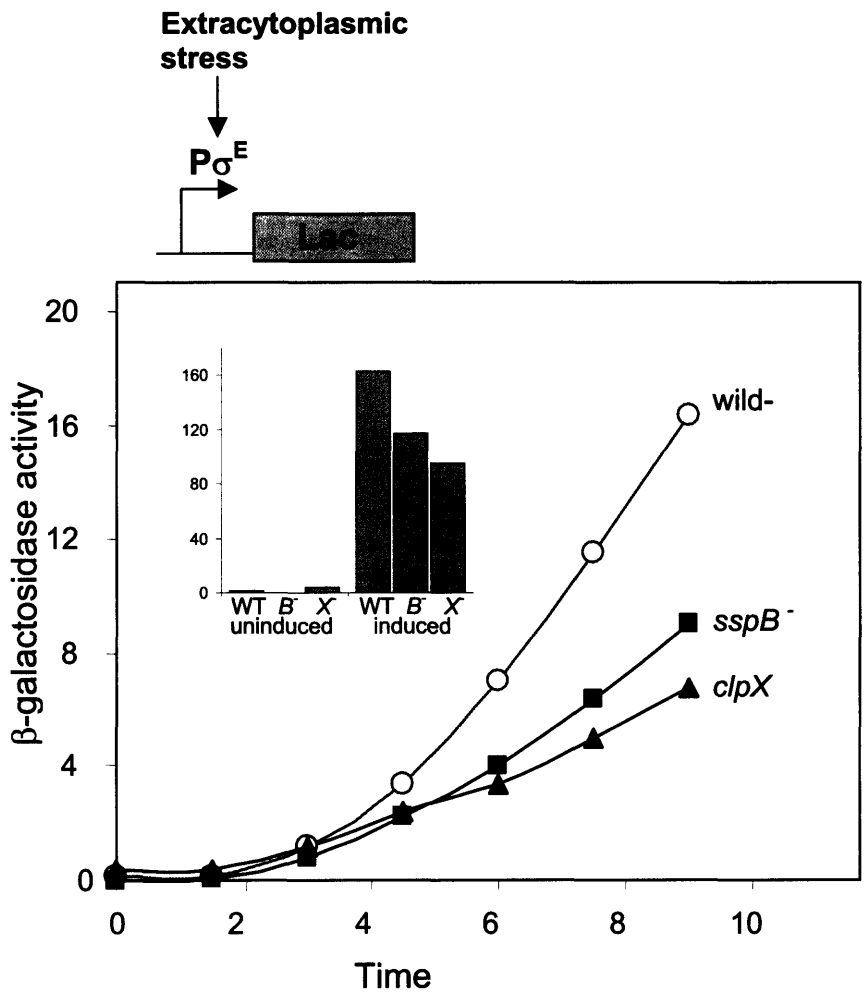


Figure 4.5. Induction of the σ^E regulon is attenuated in *sspB*⁻ and *clpX*⁻ strains.

The extracytoplasmic-stress response was induced in wild-type (CAG43583), *sspB::kan* (CAG43583) and *clpX::kan* (CAG43583) strains with L-arabinose at time zero. Samples were analyzed for β-galactosidase activity at the times indicated. The *clpX::kan* and *sspB::kan* strains grow slightly slower than wild type. When the cultures were at a similar OD_{600} however, the *clpX::kan* and *sspB::kan* strains still exhibited reduced levels of β-galactosidase (see inset). The uninduced samples were measured at time zero when the cultures were at an OD_{600} of 0.15 and the induced were measured when the cultures reached an OD_{600} of 0.45.

show that ClpX and SspB participate in activation of σ^E during the stress response. The *clpX*$^-$ cells had a larger defect than the *sspB*$^-$ cells, in accordance with the observation that SspB is not essential for ClpXP degradation of RseA$^{1-108}$ *in vitro*. Although clearly reduced, the σ^E-reporter gene was still induced in the absence of ClpX, suggesting that proteases in addition to ClpXP also participate in the activation of σ^E by degrading RseA$^{1-108}$ (see Discussion).

SspB forms stable delivery complexes with RseA$^{1-108}$ and with σ^E•RseA$^{1-108}$.

Mutagenic and crystallographic studies have identified detailed interactions between the ssrA tag and SspB and peptide-binding studies have established a strong consensus sequence for SspB recognition of the tag (Levchenko et al. 2000; Flynn et al. 2001; Levchenko et al. 2003; Song and Eck 2003). Inspection of the RseA$^{1-108}$ sequence, however, failed to identify any sequences with significant homology to the SspB-recognition sequence in the ssrA tag. Thus, we sought to determine if SspB forms a specific complex with RseA$^{1-108}$ as it does with the ssrA tag using gel filtration as a binding assay. SspB and RseA$^{1-108}$ co-eluted on a Superose 12 column at a position distinct from free RseA$^{1-108}$ (Fig. 4.6a). Moreover, a larger ternary complex was formed when SspB, σ^E, and RseA$^{1-108}$ were mixed (Fig. 4.6b). The presence of SspB, σ^E, and RseA$^{1-108}$ in this complex was confirmed by SDS-PAGE (data not shown). Stable formation of this ternary complex provides further support for the model that SspB binds the σ^E•RseA$^{1-108}$ complex and delivers this complex to ClpXP.

Truncation experiments established that a sequence near the C-terminus of RseA$^{1-108}$ was required for stable complex formation with SspB. A truncated variant ending at residue 89 (RseA$^{1-89}$) failed to co-elute with SspB during gel filtration whereas a slightly longer variant, RseA$^{1-99}$, retained the ability to bind SspB stably (Fig. 4.7a). To determine which portion of RseA$^{1-108}$ bound SspB, we looked for sites protected from tryptic cleavage in the complex.

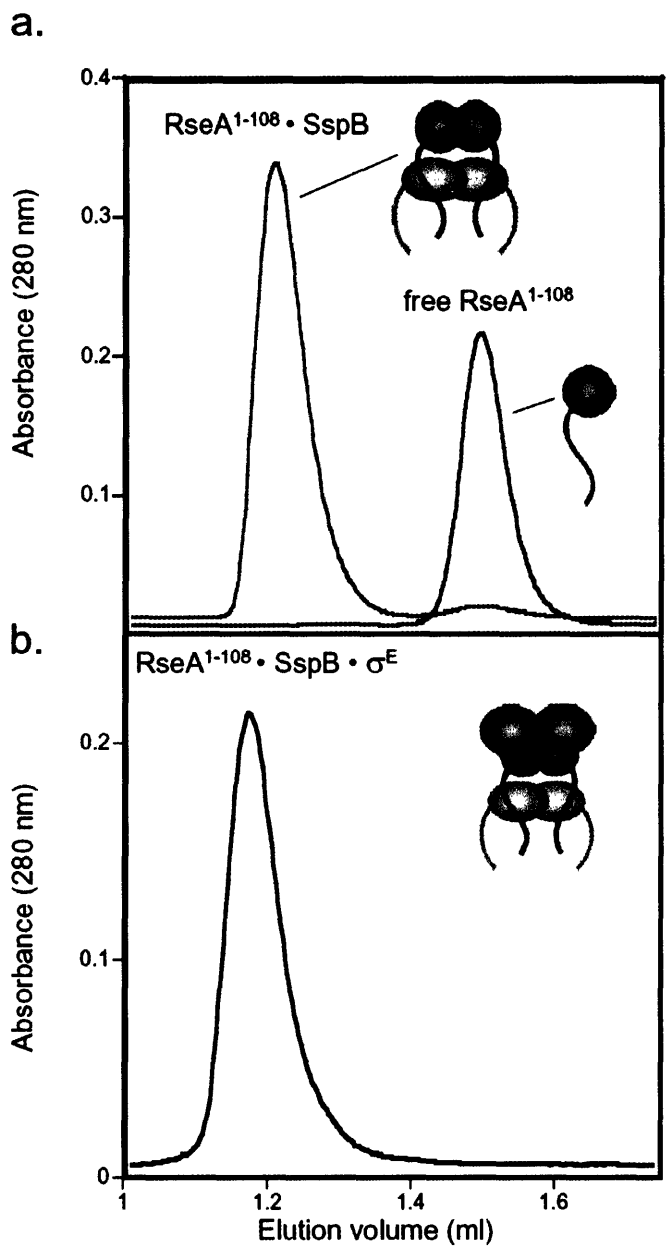
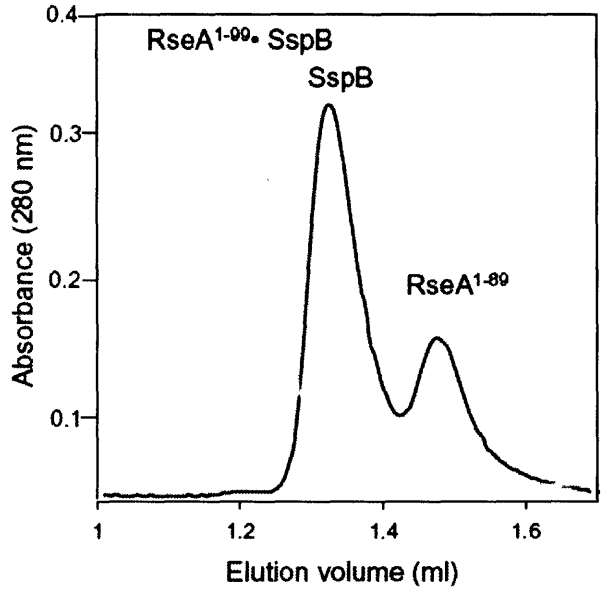


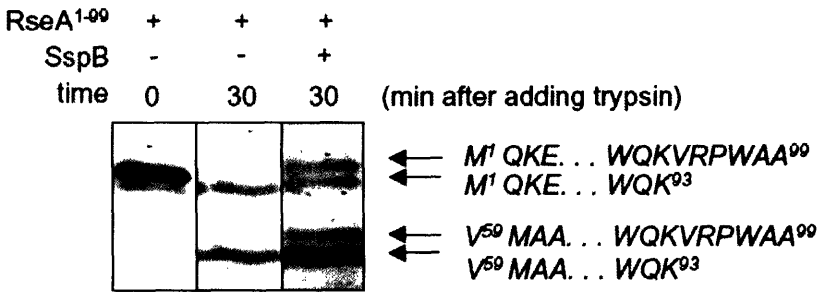
Figure 4.6. SspB forms stable complexes with $RseA^{1-108}$ and $RseA^{1-108}$•σ^E.

Gel-filtration on a Superose 12 column (4 °C) of $RseA^{1-108}$•SspB complex (panel a; top trace), free RseA (panel a; bottom trace), and $RseA^{1-108}$•σ^E•SspB complex (panel b).

a. Complex Formation



b. Trypsin Protection



c. Peptide Binding

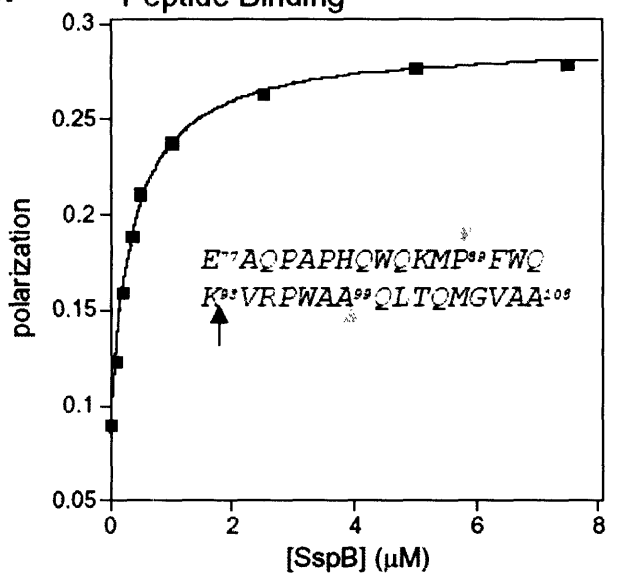


Figure 4.7. SspB interacts with residues 77-99 of RseA.

(a) $RseA^{1-99}$ forms a stable complex with SspB (gray trace), whereas $RseA^{1-89}$ does not form this complex (black trace). The RseA variants and SspB were incubated at 30 °C for 5 min and then chromatographed on a Superose 12 gel-filtration column (4 °C).

(b) Protection of $RseA^{1-99}$ by SspB from tryptic cleavage. $RseA^{1-99}$ (5 μM) was incubated with trypsin in the absence or presence of SspB (15 μM). Electrospray mass spectrometry and N-terminal sequencing determined the identity of the resulting fragments.

(c) The $RseA^{77-108}$ peptide binds to SspB. Binding of fluorescently labeled $RseA^{77-108}$ peptide to SspB at 30 °C was measured by an increase in polarization. The solid line is a fit for a K_d of 0.35 μM. Unlabeled $RseA^{77-108}$ peptide was able to compete for binding to the fluoresceinated peptide. The sequence of the $RseA^{77-108}$ peptide is given in the inset. Gray arrows correspond to the C-terminal residues of the fragments tested for complex formation in Fig. 4.7a. The black arrow corresponds to the protected trypsin site in Fig. 4.7b.

Incubation of RseA$^{1-99}$ with trypsin resulted in two major stable fragments; the larger fragment resulted from digestion after K^{93} whereas the smaller fragment was generated by trypsin digestion after both R^{58} and K^{93} (Fig. 4.7b). In the presence of SspB, two larger fragments were also observed as a result of partial suppression of the cleavage following K^{93}. These data, like the truncation experiments, implicate the sequence surrounding residue 93 in SspB•RseA complex formation.

Peptide-binding studies confirm that the C-terminal region of RseA$^{1-108}$ mediates its interaction with SspB. A synthetic fluorescein-labeled peptide containing RseA residues 77-108 bound SspB with a K_d of 0.35 μM as determined by changes in fluorescence polarization (Fig. 4.7c). This binding was competed both by excess RseA$^{1-108}$ and by an ssrA peptide (data not shown). Furthermore, a mutation in the peptide-binding cleft of SspB (Bolon et al. 2004) prevented binding of both molecules. These experiments suggest that the C-terminal region of RseA$^{1-108}$ and the ssrA peptide bind to at least some common sites within the peptide-binding cleft on SspB despite the lack of significant sequence homology.

Discussion

ClpXP and SspB regulate σ^E activity via RseA destruction.

The activity of σ^E, the transcription factor for the extracytoplasmic-stress response, is tightly controlled by its binding to and inhibition by the trans-membrane regulator, RseA (De Las Penas et al. 1997b; Missiakas et al. 1997). Stress induces sequential cleavages of RseA on each side of the membrane by the DegS and YaeL proteases, respectively, releasing the σ^E•RseA$^{1-108}$ complex into the cytoplasm (Ades 2004). Our results show that ClpXP, with the assistance of SspB, recognizes the inhibited σ^E•RseA$^{1-108}$ complex and catalyzes release of active σ^E through selective proteolytic destruction of RseA$^{1-108}$. Thus, SspB and ClpXP participate in the final stage of a proteolytic cascade, which begins in the periplasm and, ultimately, releases an active transcription factor in the cytoplasm.

ClpXP is especially well suited to recognize and degrade proteins with C-terminal signals generated by prior proteolytic cleavage. Cleavage of RseA from the membrane generates a fragment that terminates with *VAA-COOH*, a sequence that belongs to the C-motif 1 class of ClpXP recognition signals (Flynn et al. 2003). For this class of peptide sequences, which includes the ssrA tag, the non-polar side chains and the free □-carboxyl group are both important for ClpX recognition (Kim et al. 2000; Flynn et al. 2001). Thus, a *VAA* or *LAA* sequence is recognized poorly, if at all, at an internal position in a protein. Degradation of the SOS-response repressor, LexA, also illustrates this type of regulation (Neher et al. 2003a). Full-length LexA is not a ClpXP substrate, but damage-induced auto-cleavage creates an N-terminal LexA fragment, ending with *VAA-COOH*, which is degraded efficiently by ClpXP (Neher et al. 2003a). Thus, certain internal peptide sequences function as cryptic degradation signals, which remain hidden until revealed by protein cleavage. Cryptic signals permit coordinated protein destruction, allowing a single protein processing event—such as cleavage in response to an environmental cue—to trigger recognition by ClpXP.

Structural and biochemical studies demonstrate that complexes of σ^E with RseA are very stable and incompatible with transcriptional activation. The co-crystal structure of $RseA^{1-90}$ bound to σ^E reveals extensive contacts in which the first 66 amino acids of RseA are sandwiched between the two domains of σ^E in a manner that would directly block σ^E-RNA polymerase interaction (Campbell et al. 2003). We found that the σ^E•$RseA^{1-108}$ complex co-purified over several columns without detectable dissociation during a period of days (unpublished data), and direct experiments estimate the half-life of the complex *in vitro* to be well in excess of two hours (I. Grigorova and C. Gross, personal communication). Response to extracytoplasmic stress, by contrast, occurs in minutes, a time-scale similar to the rate of ClpXP degradation of $RseA^{1-108}$ in a σ^E•$RseA^{1-108}$ complex. Therefore, ClpX must actively pull the two proteins in the σ^E•$RseA^{1-108}$ complex apart to release σ^E and allow degradation of $RseA^{1-108}$. The proteolytic activity of ClpP in the ClpXP complex may assist in activation of σ^E by destroying $RseA^{1-108}$ to prevent reformation of the $RseA^{1-108}$•σ^E complex. Thus, a key feature of σ^E activation is the mechanical disassembly of the σ^E•$RseA^{1-108}$ complex by ClpXP.

In the co-crystal structure of σ^E•$RseA^{1-90}$, the first 66 residues of RseA form a stable domain that binds σ^E, while the last 24 residues are not visible and are presumably unstructured (Campbell et al 2003). Although previously there was no known function for this unstructured extension of N-RseA, our data indicates that this region functions to interact with both SspB and ClpX during the final step of activation of σ^E. SspB, σ^E and N-RseA form a stable delivery complex, in which σ^E interacts with the first 66 residues of RseA, and SspB binds to the C-terminal unstructured tail. Why is $RseA^{1-108}$ the only member of this stable complex degraded? Both SspB and σ^E probably lack degradation signals that would allow ClpX to engage these proteins to initiate protein degradation. Alternatively, the geometry of the complex might place $RseA^{1-108}$ but not the other proteins in a position that allows engagement by the enzyme.

σ^E function is essential in *E. coli* (De Las Penas et al. 1997a) but ClpX, ClpP and SspB are nonessential proteins, suggesting that other proteases also degrade $RseA^{1-108}$ and release active σ^E. Indeed, *clpX*$^-$ and *sspB*$^-$ cells show reduced induction of a σ^E-regulated promoter, rather than no induction. In fact, recent experiments demonstrate that several different proteases participate in degradation of $RseA^{1-108}$ although ClpXP plays the single largest role (R. Chaba and C. Gross, personal communication). Hence, $RseA^{1-108}$ must contain targeting signals for several proteases, emphasizing the critical nature of its destruction.

Adaptors like SspB expand and regulate the substrate repertoire of proteases.

Prior to this study, ssrA-tagged proteins were the only known substrate partners for SspB (Levchenko et al. 2000). Identification of $RseA^{1-108}$ as a new SspB partner provides the opportunity to compare mechanisms of substrate delivery. There are many similarities. Both $RseA^{1-108}$ and ssrA-tagged proteins contain a C-motif 1 degradation tag at the extreme C-terminus, and SspB binds to a nearby region within 10-30 residues. For both classes of substrates, SspB enhances ClpXP degradation principally by decreasing K_m and therefore serves to stabilize enzyme-substrate interactions. Finally, $RseA^{1-108}$ and the ssrA tag appear to occupy overlapping binding sites in the peptide-binding cleft on SspB.

Despite these similarities, the sequences within $RseA^{1-108}$ and the ssrA tag that bind SspB are not similar. Experiments presented here reveal that the SspB-binding site in $RseA^{1-108}$ lies between residues 77 and 99 (see Fig. 4.7). This region, as well as the rest of $RseA^{1-108}$, is devoid of sequences resembling the ssrA-tag consensus for SspB binding ($[AGPSV]^1$-$[ASV]^2$-$[NH]^3$-$[DCE]^4$-X^5-X^6-$[FWY]^7$) (Flynn et al. 2003). Studies are currently in progress to define more clearly how $RseA^{1-108}$ binds to SspB and how the peptide-binding cleft of SspB can interact strongly and specifically with two, seemingly unrelated, sequences.

The studies reported here revealed several different proteins that were trapped in *sspB*$^+$ but not *sspB*$^-$ strains. In addition to RseA$^{1-108}$, trapping of both AceA (isocitrate lyase) and Cdd (deoxycytidine deaminase) were also stimulated by the presence of SspB (data not shown). Delivery of ssrA-tagged substrates or RseA$^{1-108}$ for ClpXP degradation is clearly a direct consequence of SspB function, and we suspect that additional proteins will also be directly delivered by SspB. However, adaptors also can have indirect effects on substrate selection by AAA+ proteases. For example, by mediating efficient degradation of specific substrates, an adaptor may serve to free the protease to degrade other substrates more efficiently. In addition, targeted degradation of transcription factors, translation regulators, chaperones, and proteases has the potential to cause large changes in protein levels, leading to indirect changes in the repertoire of substrates available for degradation.

Although SspB is a positive regulator of RseA$^{1-108}$ recognition, it also has the potential to act as an inhibitor. In our experiments, ClpXP trapped a few substrates more efficiently when SspB was absent (see Fig. 4.1). SspB binding could prevent ClpXP degradation of certain proteins by masking their degradation tags. In fact, both SspB and the ClpS adaptor protein inhibit ClpAP recognition of ssrA-tagged proteins (Flynn et al. 2001; Dougan et al. 2002b). Alternatively, absence of competition could lead to improved degradation of substrates or substrate-adaptor complexes that compete with SspB for tethering to ClpX.

It is becoming increasingly clear that many proteins are targeted for disassembly and destruction by AAA+ ATPases both by intrinsic recognition tags and by extrinsic tethering mediated by adaptor proteins. How many adaptors exist for each enzyme, and their overall impact on recognition is not yet known. In addition to SspB, *E. coli* ClpXP uses the RssB adaptor which delivers the stationary sigma factor σ^S to ClpXP for degradation during non-starvation conditions (Muffler et al. 1996; Zhou and Gottesman 1998). Furthermore, the UmuD subunit of the UmuD UmuD' heterodimer functions as an SspB-like adaptor for UmuD' degradation by ClpXP during recovery from DNA damage (Neher et al. 2003b).

Why do certain substrates use adaptors? One answer is that adaptor proteins can increase the efficiency of recognition at low substrate concentrations. For example, SspB improves ClpXP recognition of $RseA^{1-108}$ in vivo, as shown both by trapping and σ^E-induction experiments, even though $RseA^{1-108}$ is a good ClpXP substrate in the absence of SspB in vitro. Furthermore, the use of adaptors can lead to the degradation of a group of proteins, allowing co-regulation. The results of our trapping experiments indicate that up-regulation or down-regulation of SspB would be likely to change the efficiency of degradation of a group of substrate proteins in a coordinated manner. In fact, we have observed that overproduction of SspB improves activation of σ^E during stress (data not shown). We suspect that additional adaptors remain to be discovered. These proteins, like SspB, will probably also bind a spectrum of substrates, thereby controlling the breadth and efficiency of recognition by their partner AAA+ enzymes.

Experimental Procedures

Strains and plasmids: Genes encoding RseA$^{1-108}$ and RseA$^{1-99}$ were amplified by PCR from *E. coli* genomic DNA using primers encoding NdeI and BamHI restriction sites. The amplified DNA was cleaved with both restriction enzymes and cloned between the NdeI and BamHI sites of pET3a to generate pET3a-rseA$^{1-108}$ and pET3a-rseA$^{1-99}$. A plasmid expressing RseA-DD$^{1-108}$ was constructed by site directed mutagenesis of the *rseA$^{1-108}$* gene. The gene encoding σ^E (*rpoE*) was PCR amplified from *E. coli* chromosomal DNA and cloned into the NdeI and BclI sites of the pT7LysS plasmid (IL, unpublished) to generate pT7LysS-rpoE.

The chromosomally encoded *sspB* gene was replaced by a FRT-flanked kanamycin resistance cassette following the method of (Datsenko and Wanner 2000). The *sspB::kan* cassette was then transferred into W3110 *clpP::cat ΔsmpB-1* cells by P1 transduction. Km^R mutants were transformed with pCP20 encoding the Flipase enzyme and resulting transformants were tested for loss of the kanamycin resistance as described in (Datsenko and Wanner 2000). The deletion was confirmed by PCR analysis. A *clpA::kan* cassette was then introduced by P1 transduction and finally pJF105 (Flynn et al. 2003) encoding the ClpP^{trap} was transformed into the strain (JF259). CAG43583 (Walsh et al. 2003) was a gift from Carol Gross (UCSF, San Francisco, CA). The *sspB::kan* and *clpX::kan* cassettes were introduced into the strain by P1 transduction.

Solutions: Buffer A is 50 mM Tris-HCl (pH 7.0), 50 mM NaCl, 0.5 mM DTT, 2 mM EDTA, 5% glycerol. GF buffer is 50 mM Tris-HCl (pH 7.0), 150 mM KCl, 1 mM DTT, 5% glycerol. PD buffer is as described (Kim et al. 2000).

Proteins: ClpX (Levchenko et al. 1997) and ClpP (Kim et al. 2000) were purified as described; SspB was a gift from David Wah (MIT, Cambridge, MA).

RseA$^{1-108}$ was purified from *E. coli* ER2556 pLysS/pET3a-RseA$^{1-108}$ cells grown in LB broth with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. Cells were grown at 37 °C to an OD_{600} of 0.6 and protein expression was induced with IPTG for two hours. All purification steps were performed at 4 °C. Cell pellets were resuspended at a concentration of 3 mL/g of cells in buffer A plus 6 M guanidine and Protease Inhibitor Cocktail III (Calbiochem). Following lysis for one hour, the lysate was centrifuged for 30 min at 25,000 x g and the supernatant was dialyzed overnight against 4 L buffer A with one buffer change. Insoluble proteins were removed by centrifugation, and ammonium sulfate was added to the supernatant to a final concentration of 30%. After mixing for 20 min, the precipitate containing RseA$^{1-108}$ was collected by centrifugation, resuspended in buffer A, and desalted into the same buffer using a PD-10 desalting column (Amersham Biosciences). This sample was loaded onto a MonoQ HR 5/5 column (Amersham Biosciences) equilibrated in buffer A. The column was washed with 10 column volumes of buffer A, and the bound protein was eluted with a gradient to 1 M NaCl. The peak including RseA$^{1-108}$ was collected and TFA was added to a final concentration of 0.06%. The sample was applied to a C4 HPLC column equilibrated in 0.06% TFA to separate full-length RseA$^{1-108}$ from degradation products and eluted with a gradient to 80% acetonitrile. RseA$^{1-108}$ was lyophilized, resuspended in buffer A and dialyzed against the same buffer overnight. RseA$^{1-108}$ concentration was determined by UV absorbance (ε_{280} = 24040 M^{-1} cm^{-1}).

RseA$^{1-99}$ was purified from *E. coli* BL-21 /pET3a-rseA$^{1-99}$ cells using a similar protocol except lysis in 50 mM Tris (pH 8.0), 0.I5 M NaCl, 1 mM EDTA, 5% glycerol was performed by French press and a Superdex 75 column (Amersham Biosciences) was used in place of HPLC as the final purification step. The resulting protein was greater than 95% pure as determined by Commassie staining on a SDS-PAGE gel.

The σ^E•RseA$^{1-108}$ and σ^E•RseA$^{1-108}$DD complexes were purified from ER2566 *E. coli* cells co-expressing either pET3a-rseA$^{1-108}$ or pET3a-rseA$^{1-108}$DD and pT7LysS-rpoE plasmids. The binary complexes were purified on a Sephacryl S-200 gel filtration column (Amersham Biosciences) followed by chromatography on Source15Q (Amersham Biosciences). The RseA fragment purified from the σ^E-RseA$^{1-108}$•σ^E complex contained a number of C-terminal degradation products. The smallest degradation product (RseA$^{1-89}$) was isolated by a C4 HPLC column as described above, and had a molecular weight of 10251 Da by electrospray mass spectrometry.

^{35}S-labeled RseA$^{1-108}$ and σ^E were purified as a complex from ER2566 *E. coli* cells co-expressing *pET3a-rseA$^{1-108}$* and *pT7LysS-rpoE*. ^{35}S-labeling was performed as described (Kim et al. 2000). Cells were lysed in guanidine, the lysate was dialyzed against buffer A, and a 30% ammonium sulfate cut was performed. The precipitate was resuspended in buffer A, and applied to a protein C4 HPLC column equilibrated in 0.06% TFA. RseA$^{1-108}$ and σ^E, were separated by a gradient to 80% acetonitrile, lyophilized, resuspended in buffer A, and dialyzed against the same buffer overnight. σ^E concentration was determined by UV absorbance (ε_{280} = 14650 M^{-1} cm^{-1}).

Synthetic fluorescein-labeled peptides containing residues 77-108 of RseA (*EAQPA PHQWQ KMPFW QKVRP WAAQL TQMGV AA*) and of an SsrA tag sequence (*NKKGR HGAAN DENYA LAA*) were synthesized by the MIT Biopolymers Laboratory (Cambridge, MA) and purified by reverse-phase chromatography on a C4 HPLC column (Vydac).

Protein trapping: Trapped proteins were isolated from an *sspB*$^+$ strain (JF162; W3110 *clpP::cat clpA::kan ΔsmpB-1/*pJF105) or an *sspB*$^-$ strain (JF259; see above) and analyzed by 2-D gels as described (Flynn et al. 2003). Protein spots from the gel were excised, digested with trypsin, and analyzed by microcapillary reverse-phase HPLC nano-electrospray tandem

mass spectrometry using a Finnigan LCQ DECA quadropole ion trap mass spectrometer (Harvard Microchemistry Facility). The 2-D spot corresponding to $RseA^{1-108}$ was subjected to in-gel tryptic digestion as described (Rosenfeld et al. 1992; Hellman et al. 1995) and peptides were analyzed by MALDI mass spectrometry at the MIT Biopolymers Facility (Cambridge, MA).

Degradation assays: $ClpX_6$, $ClpP_{14}$, ATP (4 mM), and an ATP regeneration system (50 μg/ml creatine kinase and 2.5 mM creatine phosphate) were mixed in PD buffer and incubated for 2 min at 30 °C. For gel analysis $RseA^{1-108}$ or $RseA\text{-}DD^{1-108}$ (2 μM) was added, and samples were removed at different times and analyzed by SDS-PAGE. Bands were visualized using Sypro Orange protein stain (Molecular Probes) on a Fluorimager 595 (Molecular Dynamics). Degradation of ^{35}S-labeled proteins were assayed by changes in TCA-soluble radioactivity as described in Burton et al. (2001). When present, the SspB concentration was 0.2 μM (monomer equivalents).

Gel filtration of protein complexes: Gel filtration was performed on a SMART system (Amersham Biosciences) using a Superose 12 column equilibrated in GF buffer at 4 °C. $RseA^{1-108}$, $RseA^{1-99}$, $RseA^{89}$ or the $RseA^{1-108}$• σ^E complex (8 μM) was incubated with or without SspB (8 μM monomer equivalents) in GF buffer for 5 min at 30 °C prior to chromatography.

Limited trypsin proteolysis: 5 μM $RseA^{1-99}$ was incubated with or without 15 μM SspB in 100 mM Tris-HCl (pH 8.9) for 5 min at 30 °C. Trypsin and $RseA^{1-108}$ were mixed in a 1:93 ratio and samples were taken at different times and analyzed by 18% Tris-Tricine SDS-PAGE. To identify the resulting RseA fragments, a portion of each time point was analyzed by

electrospray mass spectrometry and another portion was separated by SDS-PAGE, transferred onto PVDF membrane (Millipore), stained by Ponceau red stain, and subjected to N-terminal sequencing at the MIT Biopolymers Facility.

Peptide-binding assays: Binding of SspB to the fluorescein-labeled $RseA^{75-108}$ peptide (0.1 μM) was assayed by fluorescence polarization (excitation 467 nm; emission 511 nm) at 30 °C in PD buffer lacking NP-40 using a Fluoromax-2 instrument (ISA, Jobin-Yvon, Longjumeau, France). Binding curves were fit using Kaleidagraph (Synergy Software, Reading, Pennsylvania).

β-galactosidase assays: Overnight cultures were diluted 1:100 to an OD_{600} of ~0.025 and grown at 30 °C in LB broth with appropriate antibiotics. The cultures were then grown at 30 °C to an OD_{600} of 0.15 and over expression of the OmpC fusion protein was induced by 0.2% L-(+)-arabinose. β-galactosidase activities were measured as described (Miller 1972; Mecsas et al. 1993; Ades et al. 1999).

Acknowledgements

Supported by NIH grant AI-16892 and HHMI. T.A.B. is an employee of HHMI. We thank Carol Gross for strains and advice, and Rachna Chaba, Irena Grigorova and Carol Gross for sharing unpublished results. We thank the Harvard Microchemistry Facility for mass spectrometry and members of the Baker and Sauer labs for help, advice, and comments on the manuscript.

APPENDIX

I. Defining consensus Clpx-recognition motifs by screening peptide libraries

ClpX recognizes five distinct substrate binding motifs consisting of short peptide sequences (see Chapter Three). These motifs are defined based on sequence alignments of the N- and C-terminal regions of known ClpX substrates. Representative sequences from each class are sufficient to target a protein that is not normally a ClpXP substrate for degradation by ClpXP. In addition, mutational analysis has confirmed the importance of specific residues within these motifs for interactions with ClpX. For instance, mutation of the two C-terminal residues of the ssrA tag (C-motif 1) abolishes recognition of the tag by ClpX, whereas mutation of Leu^{9}, three residues before the C-terminus, decreases affinity of the tag for ClpX about 4-fold (see Chapter Two). Mutational analysis of the C-terminus of MuA (C-motif 2) indicates that the nonpolar C-terminal dipeptides and the basic side chains 3-6 residues before the C-terminus are important for ClpX recognition of MuA (Levchenko et al. 1997b). Finally, mutational analysis of N-motifs 1 and 2 has shown that a number of conserved residues within these motifs contribute to the specificity of binding to ClpX. For instance mutating residues Thr^{3}, Lys^{5}, or Leu^{6} to aspartate of a fusion protein consisting of the first 12 residues of Dps (N-motif 1) and the reporter protein Arc (Dps^{2-12}-Arc) enhances the in vitro half-lives of these proteins by an order of magnitude (see Chapter Three).

To further define the sequence rules governing substrate choice by ClpX, it is necessary to form consensus sequences for each motif based on the amino acids tolerated at each position. This information will allow us to search the *E. coli* proteome for potential ClpXP substrates based on sequence information alone. A combination of peptide library experiments similar to the ones discussed in this appendix are very promising for helping to achieve this goal.

Screening an ssrA peptide library.

Previous experiments have shown that an ssrA peptide (AANDENYALAA) with a N-terminal solubilization tag (NKKGRHG) (Sol-ssrA) is degraded by ClpXP (Kenniston et al. 2003). To identify amino acid substitutions within the ssrA-recognition motif that abrogate ClpX recognition, our experimental design was to synthesize a randomized library based on this sequence, incubate it with ClpXP, and look for members whose degradation was affected.

One experimental hurdle was to avoid altering the specificity of ClpP hydrolysis of the peptide through this mutagenesis. To minimize this problem, we inserted a strong ClpP cleavage site between the Sol and ssrA sequences, ensuring that all members of the peptide library that are able to bind to ClpX will be hydrolyzed by ClpP at this site. Although ClpP generally non-specifically cleaves peptide bonds of proteins, likely due to the high concentration of active sites within the ClpP chamber, it does appear to prefer certain cleavage sites within peptide sequences. Thompson *et al.* (1994) mapped the cleavage sites within the ClpP propeptide and found that ClpP cleaves the propeptide only between a Met and Ala sequence, preferentially with His at the P1 position. To mimic this cleavage site, we inserted a Met residue between the Sol and ssrA sequences to form the peptide: *NKKGRHMAANDENYALAA* (Sol-M-ssrA).

To map the cleavage of the Sol-M-ssrA peptide by ClpXP, we incubated the peptide with ClpXP and ATP and sequenced the degradation products by LC-MS/MS. The main degradation products were *NKKGRHMAANDENYAL* and *NKKGRHM* (data not shown), indicating that ClpP hydrolyzes the peptide mainly following the Leu^{16} and Met^{7} residues. Thus, even if mutating the C-terminus of the peptide alters ClpP hydrolysis at the C-terminal site, the peptide can still be cleaved at the internal site.

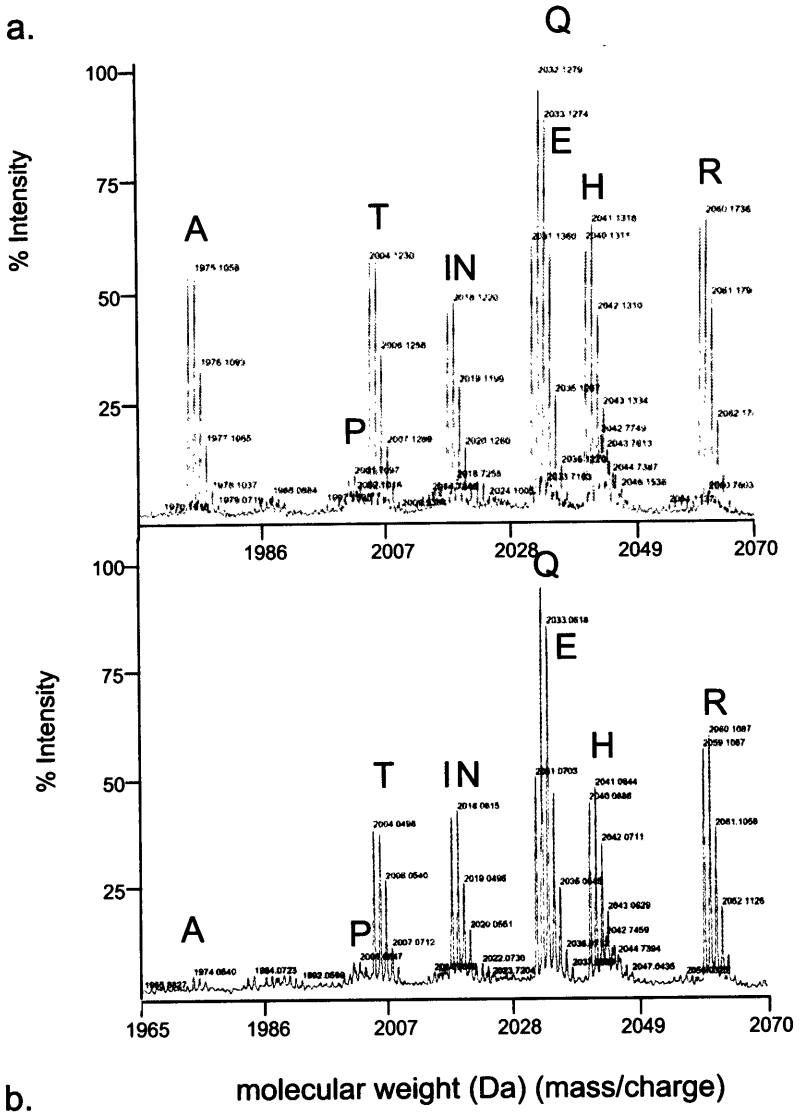
To test ClpX's sequence requirements for the C-terminal residue of the ssrA peptide, we had the following peptide library synthesized: NKKGRHMAANDENYALAX, where X was

substituted for each of the following residues: A, P, T, I, N, Q, E, H, R, W. Due to the complexity of the mass spectrometry analysis, we did not substitute this position with all 20 amino acids, and instead chose representatives with different side-chain characteristics (basic, acidic, bulky, etc.), each with a distinct molecular weight. The resulting synthesized library was under-represented in the Pro and Trp members. This library was then incubated with ClpXP and the subsequent composition of the library at various time points was analyzed by MALDI mass spectrometry.

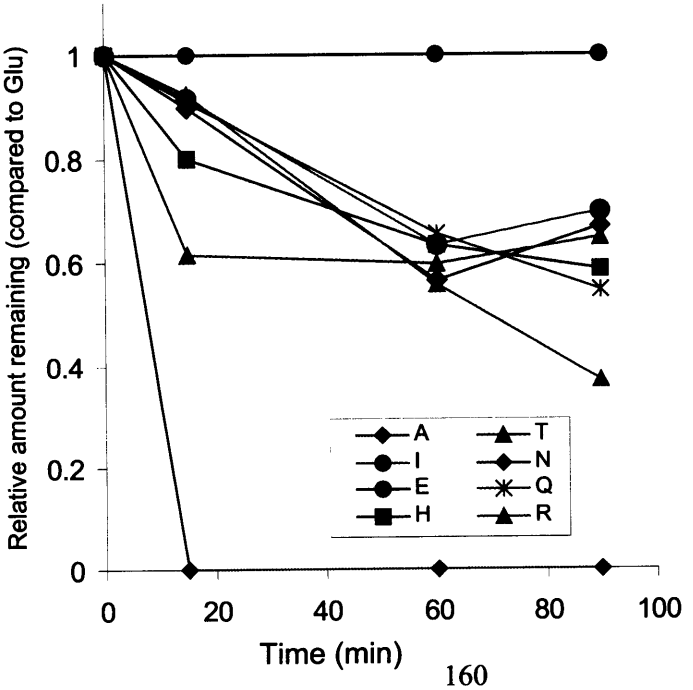
Fig. A.1 shows the chromatogram of the library composition at 0 min (upper panel) and 15 min (lower panel) following incubation with ClpXP. Based on this analysis, the peptide terminating with the wild-type residue, alanine, was degraded most thoroughly; this peptide was completely gone within 15 minutes, while a significant level of all the other library members remained. Next to alanine, the levels of threonine were reduced the most by ClpXP during this time period. This observation likely reflects the fact that small residues such as threonine and serine at this position is common within C-motif 1 (see Chapter Three).

A couple of caveats complicate the analysis of these results. First, the amount of total sample ionized by the laser varies between MALDI experiments, and thus although the relative abundance of library members within each sample likely remains constant, the total peptide concentration between samples cannot be compared. However, we know that mutating the C-terminal alanine of the ssrA peptide to aspartate completely inhibits recognition by ClpX (see Chapter Two), and by comparison, ClpX most likely does not accept the acidic residue glutamate at this site. Thus, we can make the assumption that the peptide terminating with glutamate is not degraded and normalize the peptide intensities of each sample to the intensity of the glutamate peptide peak. The graph in Fig. A.1b

a.


Q
E
H
R
A
T
IN
P
% Intensity
100
75
50
25
1986
2007
2028
2049
2070
Q
E
R
H
T
IN
A
P
% Intensity
100
75
50
25
1965
1986
2007
2028
2049
2070
molecular weight (Da) (mass/charge)

b.


Relative amount remaining (compared to Glu)
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
Time (min)
A
T
I
N
E
Q
H
R

c.

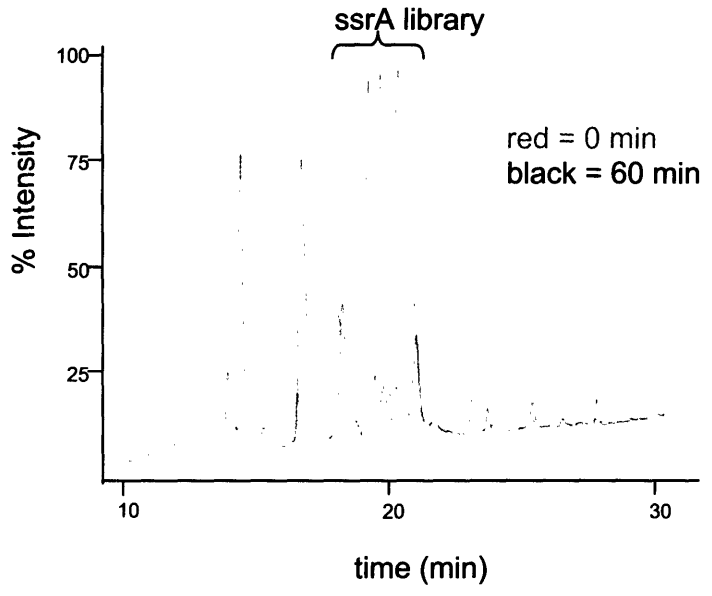


Figure A.1. Certain members of a ssrA library are degraded by ClpXP.

(a) Sol-M-ssrA (50 μM) was incubated with ClpX (0.3 μM) and ClpP (0.8 μM) and the composition of the peptide library was analyzed by MALDI mass spectrometry at 0 min (top panel) and 15 min (bottom panel).

(b) Graphic representation of data collected from experiments in (a). The intensities of each sample were normalized to the Glu peak, assuming that ClpXP does not degrade the peptide terminating in this residue. The percent of each peptide remaining compared to the 0 time point was plotted.

(c) 0 min (red) and 60 min (black) time points from the experiment in (a) were analyzed on a Protein C4 reverse phase column.

Table A.1. Calculated molecular weights of Sol-M-ssrA-X library members.

X	Molecular weight (Da)
A	1974
P	2000
T	2004
I	2016
N	2017
Q	2031
E	2032
H	2040
R	2059
W	2089

illustrates the amounts of each peptide remaining over time assuming that the peptide terminating with glutamate is resistant to degradation.

Another problem we had with analyzing the data is that each peptide has a few satellite peaks attributable to naturally occurring isotopes. Since the masses of the peptides containing the C-terminal amino acids isoleucine and asparagine, and also glutamine and glutamate only differ by one Da, it is hard to deconvolute the amount, for example, of the glutamate peak that results from glutamate itself, versus the amount due to an isotopic variant of glutamine. This only complicates the analysis of certain residues, and could likely be avoided by dividing the library into a larger number of pools so that the residues with similar molecular weights can be analyzed separately.

The evaluation of these results can be supported by analyzing each time point on a reverse phase column side-by-side with mass spectrometry. The Sol-M-ssrA-X library elutes as five distinct peaks on a C4 column. Following incubation of this library with ClpXP for one hour, three of these peaks are significantly decreased, whereas others are not greatly affected (Fig. A.1c). In addition, peaks corresponding to degradation products

appear over time. By determining which peptides contribute to each peak, it will be possible at each time point to calculate the total amount of these peptides that are degraded over time. Combining the HPLC and mass spectrometry analyses will likely be helpful in determining the rates of degradation of each member of this library.

An alternative, more expensive method would be to synthesize each member of the library separately, incubate it with ClpXP, and examine the level of degradation by HPLC. This method has become more feasible as small-scale peptide synthesis in 96 well plates has become more widely available.

The preliminary experiments performed here indicate that within the eight ssrA peptide variants tested, ClpX prefers alanine at the C-terminal residue. Next to alanine, threonine is preferred. These preferences are consistent with the amino acid variations that naturally occur within ClpXP's substrates that carry C-motif 1. This type of analysis can be applied to all of the ClpX-recognition motifs.

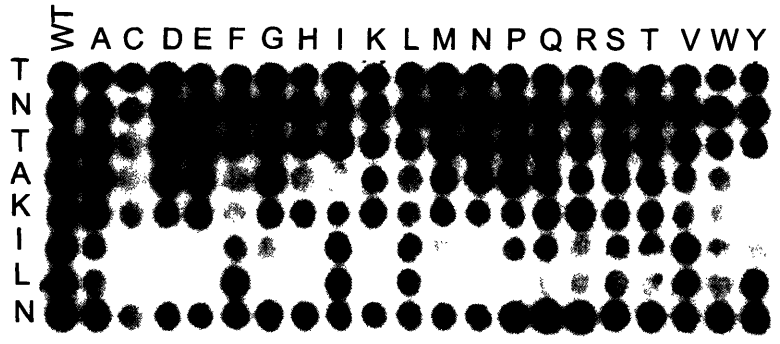
Screening of a λO peptide library.

λO contains the N-motif 1 ClpX-recognition signal: NH_2-TNTAKIL, a sequence that is both necessary and sufficient for λO's degradation by ClpXP (Gonciarz-Swiatek et al. 1999; Flynn et al. 2003). This sequence is very similar to the N-motif 1 sequence in Dps, NH_2-STAKLV; mutating the Thr^3, Lys^5, or Leu^6 residues in this motif stabilizes a Dps^{2-12}-Arc fusion protein by an order of magnitude (see Chapter Three). To further define the interaction between λO and ClpX, we synthesized an immobilized peptide library in which each of the eight N-terminal residues of λO was individually changed to each of the other 19 amino acids, whereas the rest of the sequence remained unchanged. These peptides, which contained two additional C-terminal β-alanines, were covalently attached via their C-termini to a cellulose filter by a polyethylene glycol liker.

We used a couple of methods to examine interaction of these peptides with ClpX. First, we probed the filter with non-tagged ClpX in the presence of ATPγS to allow heximerization of ClpX, however, under these conditions, no interaction between ClpX and the peptides was detected by western blot using anti-ClpX antibody. It remains unclear why this experiment did not work, but it is possible that ClpX precipitated during the incubations. As an alternative method to detect interactions, we incubated the filter with ClpX-His_6 in the absence of ATPγS and detected bound ClpX with anti-ClpX antibody. ClpX bound with a range of affinities to peptides on the filter from strong to undetectable, thus displaying a certain level of specificity (Fig. A.2a). Using an arbitrary cut-off value of 70% of wild-type binding, the following trends were observed: at residues Thr^3, Ala^4 and Lys^5, ClpX does not tolerate the bulky hydrophobic residues Phe, Trp, Tyr, Ile, Leu or Val; at Ile^6, only Val, Ile, and Leu are accepted; whereas at Leu^7 (Fig. A.2b) ClpX tolerates Phe, Trp, Tyr, Ile, Leu and Val.

These data indicate that ClpX may prefer charged or small side-chain residues at the Thr^3, Ala^4 and Lys^5 residues and bulky hydrophobic groups at Ile^6 and Leu^7. However, this trend is inconstant with the previous data indicating that mutating Thr^3 or Lys^5 to aspartate inhibits degradation of $Dps^{2\text{-}12}$-Arc. Due to the similarities of the Dps and λO recognition signals, it is unlikely that these sequences are recognized differently. However, these inconstancies could be caused by a number of different issues. First, the specificity of ClpX-His_6 as a monomer could differ from that of the untagged hexameric ClpX; a more specific binding site could be formed upon heximerization of ClpX. In addition, there is evidence that $\lambda O^{2\text{-}12}$-Arc may interact with a few different sites on ClpX (S. Siddiqui, personal communication). The XB peptide that binds specifically to the N-terminal domain of ClpX (Bolon et al. 2004) inhibits $\lambda O^{2\text{-}12}$-Arc degradation, however $ClpX^{1\text{-}46}\Delta N$ is still able to degrade $\lambda O^{2\text{-}12}$-Arc at 50% of the wild-type rate. One hypothesis is that the λO recognition

a.



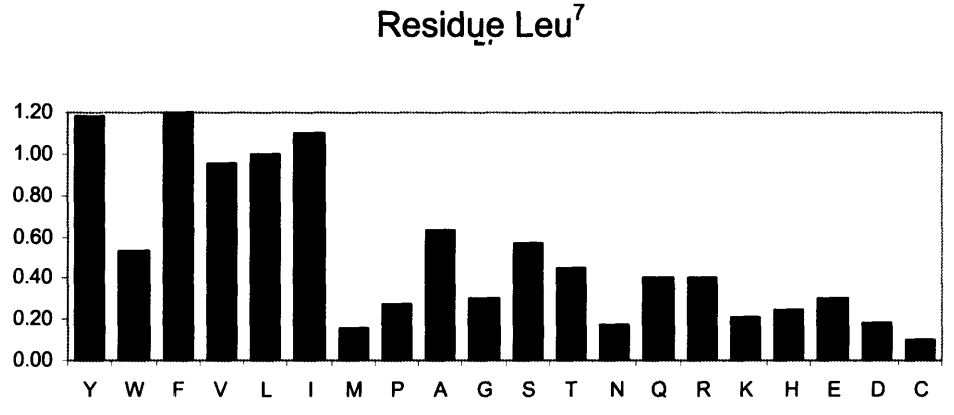


Figure A.2. Analysis of randomized λO library.

(a) A filter with covalently bound peptides corresponding to the N-terminal 11 residues of λO with each of the eight N-terminal residues randomized was incubated with ClpX-His_6 and bound protein was detected as in a western blot (see Experimental Procedures).

(b) The filter in (a) was digitally scanned and the number of pixels in each spot was quantified using ImageQuant. The data for residue Leu^7 is shown. These values are presented relative to the intensity of the wild-type λO peptide. Substitutions that show 70% or more of wild-type intensity are Tyr, Trp, Phe, Val, Leu, and Ile.

signal interacts with one site on ClpX during the initial binding step, and then a secondary site during substrate processing. Perhaps the peptide binding array experiment measures the specificity of the first initial binding interaction while the degradation of mutant proteins takes into account all of the interactions required for complete substrate processing. In this case, it is likely that a similar experiment as used for evaluating the Sol-ssrA library where degradation instead of binding is the readout, would be a more useful way of evaluating the N-motif libraries.

Experimental Procedures

Materials: Sol-M-ssrA-X library (NKKGRHMAANDENYALAX; X = A, P, T, I, N, Q, E, H, R, W) was synthesized by the MIT Biopolymers Facility (Cambridge, MA). The peptides were purified over a C4 reverse phase column (Vydac) on a Water's HPLC and the expected molecular weights were confirmed by MALDI mass spectroscopy (MIT Biopolymers Laboratory). Trp and Pro represented less than 2% of the total library. ClpX (Neher et al. 2003b), ClpX-His_6 (Levchenko et al. 1997a), and ClpP (Kim et al. 2000) were purified as described. TBS: 50 mM Tris-HCl (pH 7.5) and 125 mM NaCl. PD buffer is as described (Kim et al. 2000).

Degradation Assays: 50 μM Sol-M-ssrA-X peptide library was incubated with 0.3 μM ClpX, 0.8 μM ClpP, and ATP regeneration buffer (4 mM ATP, 50 μg/ml creatine kinase and 2.5 mM creatine phosphate) in PD buffer at 30 °C. Time points were taken as indicated and the reaction was stopped with 0.1% TFA. Part of these samples were loaded onto the HPLC C4 protein column. A linear gradient from 0.06% TFA to 80% acetonitrile, 0.06% TFA over 60 min was applied and the peptides were found to elute between 20 and 30% acetonitrile. The samples for mass spectrometry analysis were exchanged into 50% acetonitrile, 0.1% TFA using C18 ZipTips (Millipore). These samples were then analyzed by MALDI mass spectrometry at the MIT Biopolymers Facility.

Peptide array: A cellulose filter containing peptides corresponding to the 11 N-terminal residues of λO (not including the N-terminal Met residue that is removed in vivo) was prepared by the MIT Biopolymers facility using an Abimed instrument. Each peptide contained two additional C-terminal β-alanines and was covalently attached to the filter by a

polyethyleve glycol linker. The filter was soaked in ethanol, washed three times for 5 min in TBST (TBS + 0.1% Tween 20), blocked overnight in TBST plus 10% milk, washed three times with TBST, and then incubated with 3 μg/mL ClpX-His_6 in TBST plus 0.1 % milk for 6 hours overnight at 4 ˚C. The filter was then washed three times with TBST, incubated with anti-ClpX antibody for 30 min, washed three times more, and then incubated with goat-rabbit IgG HRP-conjugated antibody (Amersham) for 20 min. After three final washes, the filter was incubated with ECL substrate (NEN) and visualized on film.

II. Probing the role of the α-amino group in recognition of N-terminal ClpX–recognition motifs

The work described in this appendix was previously published as Spector S., J.M. Flynn, B. Tidor, T.A. Baker, and R.T. Sauer. *Protein Expr Purif.* **32**: 317-322 (2003). The paper is not presented here in its entirety. Sheri developed the technique to purify N-formylated proteins, and I applied it to study N-terminal recognition by ClpXP. I contributed the data for figures A2.2 and A2.3.

Abstract

Three ClpX-recognition motifs are located near the N-terminus of substrates whereas two are located at the C-terminus. The proximity of these signals to the termini could be because these are the most accessible regions of a protein and/or because the α-carboxyl and α-amino groups found only at the N- and C-terminus of a protein provide unique molecular determinants for substrate recognition. Previous experiments have shown that ClpX uses the free α-carboxyl group as a determinant for recognition of the C-motif 1 signal (Kim et al. 2000). For the N-terminal ClpXP degradation tags, it is not known whether the free α-amino group is required for recognition. The N-terminal methionine itself is a recognition determinant for one of the N-motif signals, N-motif 2. However, N-motif 1 and N-motif 3 signals can be located as far as four residues from the N-terminus. To test the requirement of the α-amino group as a ClpX-recognition determinant, we purified an N-formylated form of a ClpXP substrate and tested its degradation by ClpXP. We showed that this amino group is not required for ClpXP-mediated degradation of proteins bearing this N-terminal recognition signal.

Introduction and Results

Purification of N-formylated proteins.

In bacteria, protein synthesis initiates with formyl-methionine (fMet) (Dixon 1972). As schematized in Fig. A.3 the formyl group is then removed post-translationally by peptide deformylase (PDF), leaving a free α-amino group that is positively charged at neutral pH. Depending on the identity of the second amino acid in the protein chain, deformylation may be followed by removal of the initiating methionine by the enzyme methionine aminopeptidase (MAP). Actinonin is a PDF inhibitor (Chen et al. 2000) but is normally ineffective in *Escherichia coli* because it is removed from the cell by efflux pumps involved in multidrug resistance. However, the antibiotic kills *E. coli* strains bearing a deletion of the *acrAB* efflux pump genes (Chen et al. 2000). Thus, N-formylated proteins can be expressed in high yield in E. coli Δ*acrAB* strains if actinonin is added at the time of induction of protein expression. This system is useful to probe the function of the N-terminal α-amino group without altering the identity or position of the N-terminal amino acid. We used this technique to probe the necessity of this group as a ClpX-recognition determinant in the N-motif 2 class of signals.

Purification of fMet-IscS$^{1-11}$–Arc.

Arc repressor is not normally a substrate for the ClpXP protease; however, fusion of Arc repressor to the N-terminal 11 residues of IscS, a cysteine desulfurase (IscS$^{1-11}$–Arc) targets this fusion protein for degradation by ClpXP (see Chapter Three). IscS contains an N-motif 2 ClpX-recognition signal, a class of signals with the consensus NH_2–Met–Lys–Φ–Φ–X_5–Φ (Φ=nonpolar). To determine whether a free N-terminus is required for its degradation, IscS$^{1-11}$–Arc was expressed from an overproducing plasmid in the Δ*acrAB E. coli* strain AG100A DE3) in the presence or absence of actinonin. Following a single Ni–NTA affinity

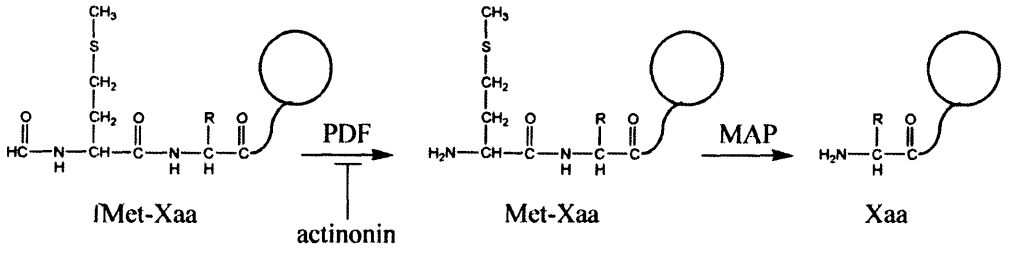


Figure A.3. Processing of newly synthesized proteins in bacteria.

Translation initiates with N-formyl methionine. Once translation is complete, peptide deformylase (PDF) removes the formyl group from fMet. Actinonin inhibits PDF, blocking this step of processing. Depending on the identity of the second amino-acid in the protein sequence, the methionine residue may be removed by methionine aminopeptidase (MAP), but this step is contingent on the removal of the formyl group by PDF.

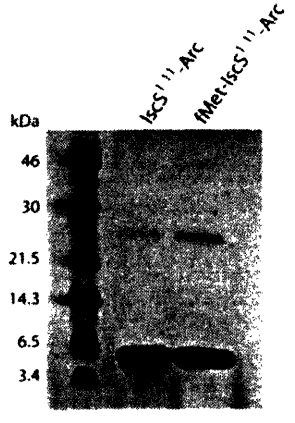


Figure A.4. SDS–PAGE analysis of protein purity

SDS–PAGE shows that $IscS^{1-11}$–Arc and fMet–$IscS^{1-11}$–Arc are >95% pure. The only observable impurity at a molecular weight of approximately 25 kDa corresponds to SlyD, a histidine-rich *E. coli* protein which often co-purifies on Ni–NTA resin with His-tagged protein.

chromatography step, these proteins were greater than 95% pure as assayed by SDS–PAGE (Fig. A.4). The wild-type protein had a mass of 9114 Da (9111 Da calculated) and the protein expressed in the presence of actinonin had a mass (9142 Da observed and 9139 Da calculated) that is consistent with retention of the fMet to produce fMet–IscS$^{1-11}$–Arc.

Degradation of fMet–IscS$^{1-11}$–Arc by ClpXP.

IscS$^{1-11}$–Arc and fMet–IscS$^{1-11}$–Arc were tested for the ability to be degraded by ClpXP. In each case, ClpXP was briefly incubated with ATP and an ATP regeneration system, protein substrates were added, and aliquots were removed after various times for analysis by SDS–PAGE. As shown in Fig. A.5, IscS$^{1-11}$–Arc and fMet–IscS$^{1-11}$–Arc were degraded at very similar rates, indicating that the N-terminal amino group of the fusion protein is not required for degradation by ClpXP.

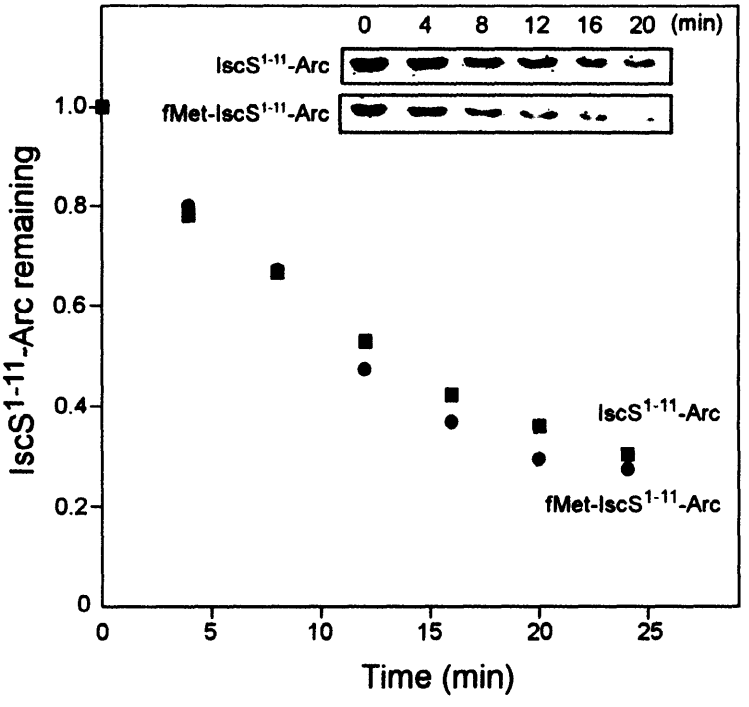


Figure A.5. ClpXP protease assays.

$IscS^{1-11}$–Arc and fMet–$IscS^{1-11}$–Arc were degraded by ClpXP protease. The fraction of substrate remaining was determined by SDS–PAGE (inset) and is plotted as a function of time.

Discussion

Expression in the presence of actinonin in an *acrAB* deletion strain of *E. coli* provides a simple method for obtaining proteins with modified N-termini. This, in turn, provides a straightforward way to test the role of the free α-amino group in systems in which it appears important for protein function. In the studies described here, we purified fMet–IscS$^{1-11}$–Arc and tested whether blockage of the α-amino group affected degradation by ClpXP. The N-terminal residues of *E. coli* IscS target an Arc fusion protein for ClpXP degradation. Moreover, a number of other ClpXP substrates, like IscS, share the consensus NH_2–Met–Lys–Φ–Φ–X_5–Φ (Φ=nonpolar). It seemed possible therefore that the free α-amino group in these proteins represented a recognition determinant for ClpXP. This does not, however, appear to be the case. We found that fMet–IscS$^{1-11}$–Arc was degraded at the same rate as Met–IscS$^{1-11}$–Arc (Fig A.5). Because the N-terminal amino group is not required for ClpXP binding or degradation, it will be interesting to determine whether the sequence motif shared by this group of ClpXP substrates could target proteins for degradation at an exposed internal or even a C-terminal position in a protein sequence.

Experimental Procedures

Plasmids and strains: AG100A (*E. coli* K-12 Δ*acrAB*) was a generous gift from Nikaido and Levy (Okusu et al. 1996; White et al. 1997). To enable expression from pET vectors, this strain was transduced with DE3 (Novagen, Madison, WI). A plasmid derived from pET11a encodes a fusion protein $IscS^{1-11}$–Arc consisting of the first 11 residues of IscS (MKLPIYLDYSA) followed by Arc-st11 (see Chapter Three). This gene encoded a C-terminal His_6 tag for Ni–NTA purification.

Expression and Purification: AG100A(DE3) cells were transformed with the pET11a-$IscS^{1-11}$–Arc plasmid and plated on LB agar with 100 μg/ml ampicillin. A single colony was picked and grown overnight at 37 °C in LB plus 100 μg/ml ampicillin and the overnight culture was diluted to prepare a 1 L culture for growth and induction under the same conditions. Cells were grown to an OD_{600} of 0.6. Expression was induced either by addition of 1 mM IPTG or 1 mM IPTG plus actinonin at a final concentration of 2 μg/ml. This actinonin concentration is 8-fold higher than the minimum inhibitory concentration, defined as the minimum concentration required such that no culture growth is observed after 18–24 h at 35 °C, as measured for AG100A (Chen et al. 2000). After 2 h, cultures were harvested by centrifugation in a Beckman J-6B centrifuge at 4000 rpm, 4 °C, for 10 min. Cell pellets were stored at −80 °C prior to lysis and protein purification.

$IscS^{1-11}$–Arc was C-terminally His-tagged and purified by Ni–NTA affinity chromatography using the standard protocol for Arc-st11 (Milla et al. 1993). To purify the wild-type and formylated protein, the cell pellet was resuspended in a pH 8 buffer containing 0.1 M NaH_2PO_4, 0.01 M Tris, 6 M guanidine hydrochloride, and 10 mM imidazole. Cells were lysed by French Press. After centrifugation, the supernatant was applied to Ni–NTA resin (Qiagen, Valencia, CA). The column was washed extensively with the above buffer and protein was

eluted in 0.2 M acetic acid, 6 M guanidine hydrochloride. After elution, fMet–IscS$^{1-11}$–Arc was dialyzed into 50 mM Tris, 250 mM KCl, 0.1 mM EDTA, pH 7.5. These proteins were greater than 95% pure as assayed by SDS–PAGE (Fig. 3.2).

Degradation assays: Degradation reactions were performed as follows: $ClpX_6$ (0.3 μM), $ClpP_{14}$ (0.8 μM), ATP (4 mM), and an ATP regeneration system (50 μg/ml creatine kinase and 2.5 mM creatine phosphate) were mixed in PD buffer (25 mM Hepes–KOH, pH 7.6, 5 mM $MgCl_2$, 5 mM KCl, 15 mM NaCl, 0.032% (v/v) Nonidet P-40, 10% (v/v) glycerol) and incubated for 2 min at 30 °C. The protein substrate (5 μM) was added and samples were removed at specific times and analyzed by SDS–PAGE.

APPENDIX III: Identification of substrates that may depend on RssB to be targeted to ClpXP for degradation

RssB (Regulator of Sigma S) was originally identified as a protein necessary for the rapid degradation of σ^S (Bearson et al. 1996; Muffler et al. 1996; Pratt and Silhavy 1996). During exponential phase growth, RssB~P binds to a motif, KExxVY, known as the destruction box, in the interior of σ^S, and facilitates its degradation by ClpXP (Bouche et al. 1998; Becker et al. 1999). RssB is dephosphorylated during stationary phase and thus loses its ability to interact with σ^S (Bouche et al. 1998). Thus, σ^S is stabilized against ClpXP degradation and is available to bind to core RNA polymerase and activate the expression of genes required to respond to this stress.

Is σ^S the only target for this adaptor protein? It is attractive to consider that there are other proteins whose correct temporal degradation is regulated by RssB. A group of ClpXP substrates are proteins that are active during stationary phase (see Chapter Three) and these would be good candidates for RssB-dependent substrates. As described in Chapter Four, we found that another well-characterized adaptor protein, SspB, is indeed able to deliver multiple partners to ClpXP. This is a powerful strategy used by ClpXP to expand the repertoire of proteins it can recognize. To look for additional RssB-dependent substrates, we captured proteins in $ClpXP^{trap}$ in strains containing or lacking RssB, and looked for proteins whose presence was higher in the trap purified from the *rssB*$^+$ strain.

RssB influences recognition of a set of ClpXP substrates in vivo.

Proteins that co-purified with $ClpP^{trap}$ in *rssB*$^+$ (Fig. A.6a) or *rssB*$^-$ (Fig. A.6b) strains were visualized by staining of 2-D electrophoresis gels. The amount of σ^S trapped in the *rssB*$^-$ strain was ten-fold lower than that trapped in the *wild-type* strain. This is consistent with results indicating that σ^S degradation by ClpXP in vitro is stimulated more than 10-fold by

RssB (Zhou et al. 2001) and validates the ability of this method to identify RssB-dependent substrates. The majority of proteins are present at similar levels within the trap purified from both strains, indicating that, as expected, most proteins do not require RssB to interact with ClpXP. However, closer inspection reveals that a handful of proteins are more abundant in $ClpP^{trap}$ when RssB is present in the cells. This indicates that, similar to SspB (see Chapter Four), RssB influences the degradation by ClpXP of substrates other than σ^S and may participate directly in delivery of these substrates to ClpXP.

2-D gel spots whose intensities were higher on the *rssB*$^+$ gel were excised, subjected to trypsin proteolysis, and analyzed by tandem mass spectrometry at the Harvard Microchemistry Department. Two spots were identified – RplJ (Ribosomal protein L10) and NrdH (Glutadoxin-like Protein).

L10 was one of the most abundantly trapped proteins originally captured by $ClpXP^{trap}$ (see Chapter Three) and has been previously characterized as an unstable protein in vivo (Petersen 1990). Like a number of ribosomal proteins, L10 is a translational inhibitor of its own operon which encodes L10, L7/L12 and the RNA polymerase β and β' subunits. Rapid synthesis of ribosomal subunits is required during exponential growth in nutrient rich medium; degradation of uncomplexed L10 during this growth phase would relieve repression of the operon, enhancing synthesis of ribosomal subunits. In contrast, synthesis of ribosomes must be slowed during stationary phase when amino acid supplies are scarce (Davis et al. 1986). An adaptor protein that controls the growth-phase dependent degradation of L10 would be one mechanism to control expression of its operon. In addition, L10 contains a sequence near its N-terminus, RExxVY, which resembles the σ^S destruction box. These facts make L10 an attractive candidate as an RssB-dependent ClpXP substrate.

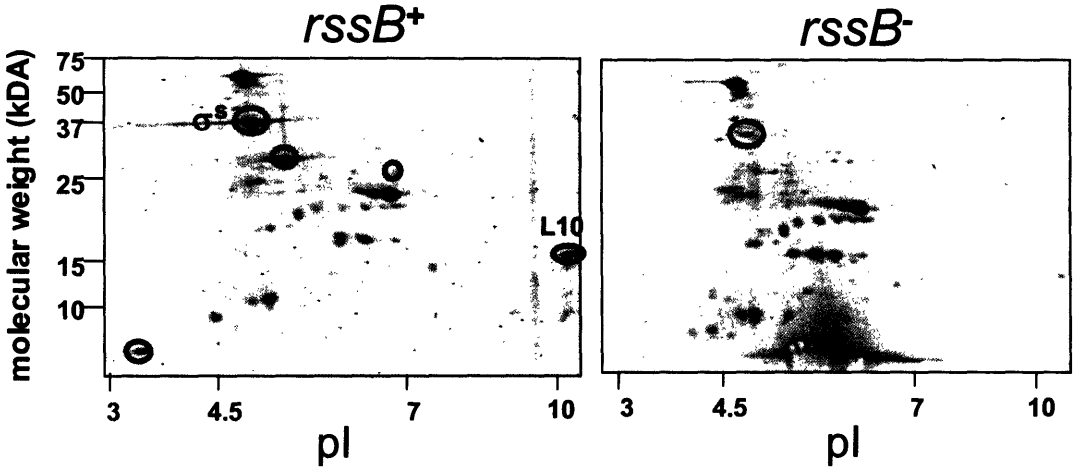


Figure A.6. 2-D gels of proteins captured by $ClpP^{trap}$ in *rssB*$^+$ and *rssB*$^-$ strains.

Gels show proteins captured by $ClpP^{trap}$ in *E. coli* strains JF169 (*rssB*$^+$; left panel) and JF243 (*rssB*$^-$; bottom panel). σ^S is circled in purple. Additional proteins whose levels are higher in the *rssB*$^+$ strain are circled in pink.

Uncomplexed L10 and NrdH are in vitro substrates for ClpXP.

Previous studies have shown that L10 is rapidly degraded when overexpressed on a plasmid; however, it is significantly stabilized upon co-expression of its binding partner L7/L12 (Petersen 1990). We purified L10 from the ribosomal fraction of cells. It is not possible to highly over-express L10 on a plasmid because of its toxicity. Purified L10 is rapidly degraded by ClpXP with an in vitro half-life under our standard laboratory conditions of about one minute (Fig. A.7, left panel). This rate of degradation is comparable to the in vivo rate. L10 carries a C-motif 1 ClpX-recognition sequence of Ala-Ala that most likely targets it to ClpXP for degradation (see Chapter Three).

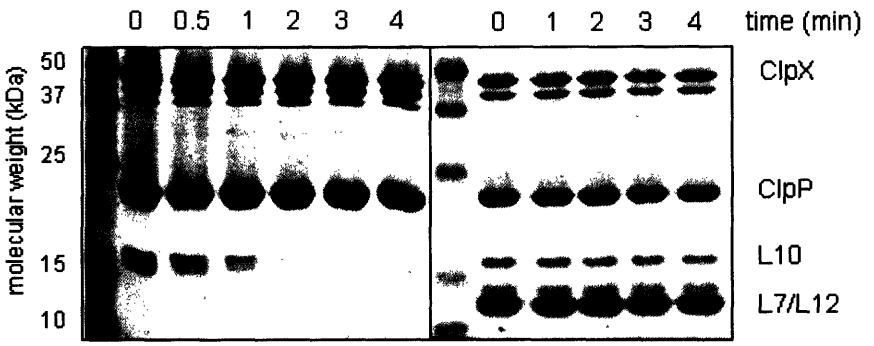


Figure A.7. L10 is rapidly degraded by ClpXP in vitro and stabilized when in complex with L7/L12.

Left panel: Degradation of 2 μM L10 by 0.3 μM $ClpX_6$ and 0.8 μM $ClpP_{14}$ was assayed by SDS PAGE gel followed by staining with Sypro Orange Stain (Molecular Probes).
Right panel: Prior to addition to ClpXP as above, 2 μM L10 was incubated with 4 μM L7/L12 at 30°C for 5 min.

L10 forms a complex with the L7/L12 subunits; this complex is incorporated into the 50S ribosome and forms the ribosomal stalk. This complex is resistant to degradation in vivo (Petersen 1990). We found that complex formation of L10 with L7/L12 inhibits its degradation by ClpXP in vitro (Fig. A.7, right panel). Structural and mutational studies performed on the L10/L7/L12 complex show that L7/L12 binds to the C-terminal region of L10 (Griaznova and Traut 2000). These data indicate that L7/L12 masks the L10 C-motif 1 recognition signal from interaction with ClpXP, allowing the subunit to be stably incorporated into the 50S ribosome.

NrdH is a glutathione-like redoxin protein. We expressed and purified a His-tagged version of NrdH. NrdH has N-motif 3 and C-motif 1 ClpX-recognition signals. ClpXP also was

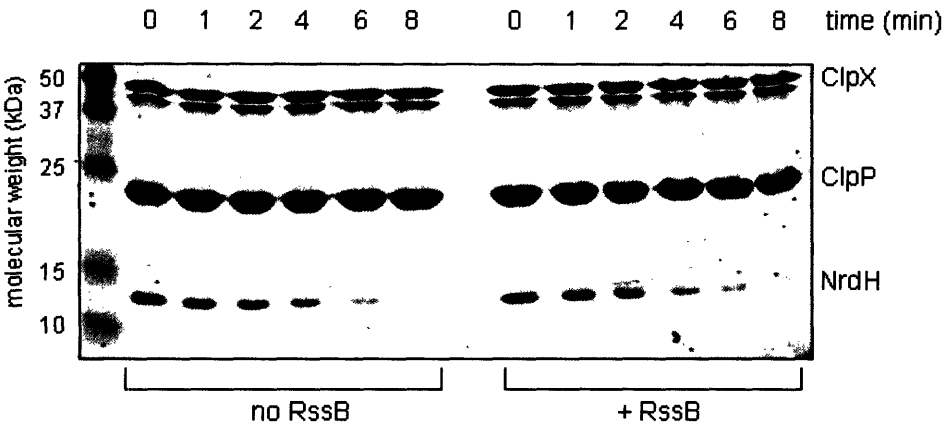


Figure A3.3. Degradation of NrdH is not activated by RssB.

Left panel: Degradation of 2 μM His_6-NrdH by 0.3 μM $ClpX_6$ and 0.8 μM $ClpP_{14}$ was assayed by SDS PAGE with (left panel) or without (right panel) 0.1 μM RssB.

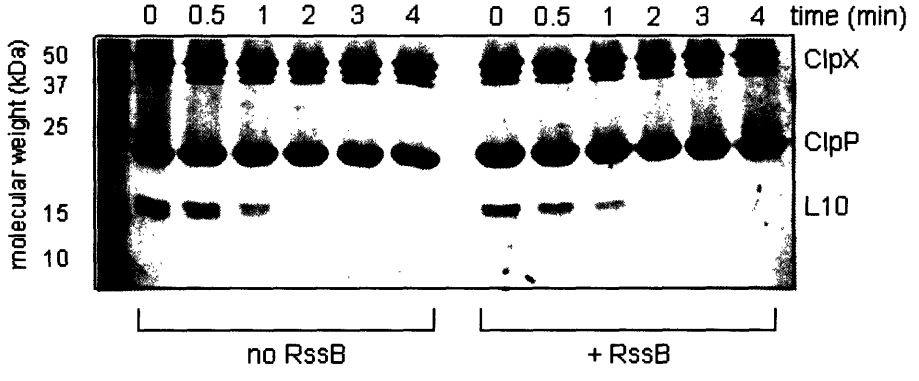


Figure A.9. Degradation of L10 is not activated by RssB.

Degradation of 2 μM L10 by 0.3 μM $ClpX_6$ and 0.8 μM $ClpP_{14}$ was assayed by SDS PAGE gel with (left panel) or without (right panel) 0.1 μM RssB~P.

able to rapidly degrade this protein in vitro in an ATP-dependent manner (Fig. A.8, left panel and data not shown).

In vitro degradation of L10 and NrdH does not appear to be stimulated by RssB.

RssB does not appear to activate the degradation of L10 or NrdH; in these experiments, we used the same in vitro conditions under which σ^S is rapidly degraded by ClpXP in a RssB-dependent manner (data not shown; Fig. A.8, Fig. A.9).

There are a number of issues that could explain the difference between the in vivo trapping results and these in vitro degradation results. First, the *rssB*$^+$ and *rssB*$^-$ strains are genetically different. Although both of the trapping strains are *clpP*$^-$, they are both *clpP*$^{trap+}$. Thus, although σ^S is not degraded by ClpXP in either strain, part of the σ^S population may be sequestered inside ClpP^{trap} in the *rssB*$^+$ strain. In contrast, in the *rssB*$^-$ strain, σ^S is not targeted to ClpXP^{trap} and thus, more σ^S is available to interact with RNA polymerase and modulate gene expression. Thus, these two strains may be expressing different levels of certain proteins and this may change the repertoire of proteins available for capture by ClpP^{trap}.

Another reason we may not see the effect of RssB on degradation of NrdH and L10 in vitro is that we may not have the right in vitro degradation conditions to measure this effect. Unlike σ^S, L10 and NrdH are rapidly degraded by ClpXP in vitro in the absence of RssB. Perhaps altering the in vitro degradation conditions to reduce the affinity of these substrates for ClpX would change the dependence on RssB for degradation. The role of these adaptor proteins in targeting purified and highly concentrated substrates to ClpXP in vitro could be very different than their roles in the competitive environment of the cell. In the cell, there are many substrates that are competing for a small number of ClpXP molecules (C. Farrell, unpublished data). In addition, there are at least three, and likely a number more, adaptor proteins that compete for binding to the N-terminal domain of ClpX (Neher et al. 2003b; Bolon

et al. 2004; Siddiqui 2004). It is possible that the role of RssB in this complex environment cannot be reproduced under our standard in vitro degradation conditions. However, it is likely that if RssB plays a role in tethering L10 or NrdH to ClpX, that these proteins directly interact. Looking for complex formation between L10 or NrdH and RssB by gel filtration would be a good starting point in determining the potential for these proteins to interact.

L10 and NrdH are two newly characterized ClpXP substrates. Both substrates have C-motif 1 recognition signals, emphasizing the role of this class of sequences in substrate recognition by ClpXP. Further work will have to be performed to validate the role of RssB in targeting these substrates to ClpXP for degradation.

Experimental Procedures

Strains and plasmids: The chromosomally encoded *rssB* gene was replaced by a FRT-flanked kanamycin resistance cassette following the method of Wanner et al. (2002). The *rssB::kan* cassette was then transferred into a W3110 *clpP::cat ΔsmpB-1* background by P1 transduction. Km^R mutants were transformed with pCP20 encoding the Flipase enzyme and resulting transformants were tested for loss of the kanamycin resistance as described in Wanner et al (2000). The deletion was confirmed by PCR analysis. A *clpA::kan* cassette was then introduced by P1 transduction and finally pJF105 (Flynn et al. 2003) encoding the $ClpP^{trap}$ was transformed into the strain (JF243).

A gene encoding NrdH was amplified by PCR from *E. coli* genomic DNA using primers encoding NdeI and BamHI restriction sites. The amplified DNA was cleaved with both restriction enzymes and cloned between the NdeI and BamHI sites of pET28b. The resulting protein had a N-terminal His_6 purification tag.

Materials: ClpX (Levchenko et al. 1997a) and ClpP-His_6 (Kim et al. 2000) were purified as described. His_6-NrdH was purified using the same method as ClpP-His_6 (Kim et al. 2000) and then desalted into 50 mM Tris, pH 7.5, 250 mM KCl, 10% glycerol. RssB and σ^S were kindly provided by S. Wickner (NIH, Bethesda, Maryland).

The L8 complex (L7/L12/L10) was purified from intact ribosomes according to Uchiumi et al. (1999) with the following modifications: Salt-washed 70S ribosomes were purified as follows: 2 L of W3110 cells were grown in LB at 37°C to an OD_{600} = 0.7. Cells were harvested and lysed by French Press. The ribosomes were extracted with 20 mL buffer A (10 mM $MgCl_2$, 20 mM NH_4Cl, 5 mM 2-mercaptoethanol, and 20 mM Tris-HCl, pH 7.5). The lysate was centrifuged twice for 30 min at 20,000 x g. The ribosomes were then pelleted at 21,000 rpm for 13 hrs in a Ti50.2 rotor at 4°C. The pellet was then resuspended in 40 mL buffer B (10

mM $MgCl_2$, 0.5 M NH_4Cl, 5 mM 2-mercaptoethanol, 20 mM Tris pH 7.5). Ribosomes were pelleted by ultracentrifugation for 3 hrs at 45,000 rpm in the same rotor. The salt wash was repeated two more times. The final pellet was resuspended in 1 ml extraction buffer (20 mM $MgCl_2$, 1 M NH_4Cl, 10 mM 2-mercaptoethanol, 40 mM Tris pH 7.5) and stored at -20°C. The L8 complex (L7/L12/L10) was removed as follows: The salt-washed ribosomes in extraction buffer were pre-incubated at 30°C for 5 min. The solution was mixed with 0.5 ml pre-warmed ethanol with stirring at 30°C. After 10 min, another 0.5 ml of ethanol was added, and stirring was continued for 5 min at 30°C. The solution was centrifuged at 15,000 rpm in an SS34 rotor for 10 min. The ribosomes were then resuspended in 2 mL extraction buffer and precipitated with 80% ice cold acetone for 3 hrs on dry ice. The precipitated ribosomes were spun down at 15,000 rpm in the SS34 rotor for 10 min and then resuspended in 6 M urea to separate the complex. The sample was filtered through a 0.45 μ spin filter (Corning) and applied to a protein C4 HPLC column equilibrated in 0.06% TFA. L10 and L7/L12 were eluted from the column using the following non-linear gradient: (0 to 40% B for 10 min; 40-90% B for 45 min; 90-100% B for 5 min; B = 0.06% TFA, 80% acetonitrile). These proteins were lypophilized, and L7/L12 was resuspended in storage buffer (50 mM Tris pH 7.5, 100 mM KCl, 0.2 mM EDTA). L10 was resuspended in 100 μl 6 M urea, 20 mM Tris pH 9.0, 0.2 mM EDTA and applied to a MonoQ 5/5 (Amersham Pharmacia) column equilibrated in the same buffer. L10 was eluted with 1 M KCl in the same buffer, and dialyzed into storage buffer.

Degradation In Vitro: 0.3 μM $ClpX_6$, 0.8 μM $ClpP_{14}$, ATP (4 mM), and an ATP regeneration system (50 μg/ml creatine kinase and 2.5 mM creatine phosphate) were mixed in SD buffer (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 140 mM KCl, 1 mM DTT, 0.1 mM EDTA, and 5% glycerol) and incubated for 2 min at 30°C. When indicated, 0.1 μM RssB and 50 mM acetyl phosphate were added following this incubation. Substrates were added at the concentrations indicated

and samples were removed at specific times and analyzed by SDS-PAGE. Bands were visualized using Sypro Orange protein stain (Molecular Probes) on a Fluorimager 595 (Molecular Dynamics).

Protein Trapping In Vivo: Trapped proteins were isolated from an *rssB*$^{+}$ strain (JF162; W3110 *clpP::cat clpA::kan ΔsmpB-1/pJF105*) and an *rssB*$^{-}$ strain (JF243; see above) as previously described (Flynn et al. 2003). Samples for 2D gel analysis were prepared and analyzed as described (Flynn et al. 2003).

Mass spectrometry: Samples for MS/MS analysis were separated by 2D gel electrophoresis as described (Flynn et al. 2003). Gel spots were excised, digested with trypsin, and analyzed by microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry using a Finnigan LCQ DECA quadropole ion trap mass spectrometer (Harvard Microchemistry Facility).

CHAPTER FIVE:

Discussion

(The Tail's End)

The work described in this thesis has given new insights into the biological roles of ClpXP and provided a foundation for understanding the strategies this protease uses to select its targets for degradation. Identification of substrates, definition of substrate binding motifs, and characterization of adaptor proteins are just a few of the recent advances that have illuminated this field. However, many areas of study remain unnavigated and many questions remain unanswered. This section poses a number of questions that may define the future directions in the area of regulated target selection by ClpXP.

What is the complete set of sequence rules that governs ClpX-substrate specificity?

Our studies have brought us closer than ever to defining the primary ClpX-interaction motifs. These sequences have been categorized into five classes. Greater than 50 new ClpXP substrates have been identified, and majority of these substrates contain sequence motifs that that are attractive candidates for ClpX-recognition signals. The C-motif 1 signal is present in almost half of all ClpXP substrates. From the culmination of numerous experiments, we can now predict with high certainty that if a protein terminates in Ala-Ala or other variants of a C-motif 1 signal and this signal is structurally available, this protein will be a substrate for ClpXP. However, our ultimate desire is to fully define ClpX's binding preferences for the five different motifs, and use this sequence information to systematically identify ClpXP substrates. To further define these motifs, we can use techniques such as those described in Appendix I. For example, we can probe degenerate libraries to identify sequences sufficient for binding by ClpX or degradation by ClpXP. Once these sequence rules governing substrate recognition by ClpXP have been fully defined, bioinformatic techniques can be applied to search for new substrates.

Following the identification of many new ClpXP substrates, we attempted to describe the similarity of their N- and C-terminal sequences using substitution matrices such as BLOSUM62. These matrices are derived from sequence alignments of protein homologs and

mainly reflect acceptable amino acid substitutions that don't severely affect protein structure (Henikoff and Henikoff 1992). However, these programs did not perform well in aligning our substrates, likely because the forces that define protein structure are in some cases different from those important for protein-protein interactions. For instance, at a binding site it may be acceptable to substitute a polar residue whose main contribution comes from its aliphatic side-chain with a non-polar residue of roughly the same size. In addition, a charge reversal such as glutamate to arginine is common in protein sequences, but could likely abolish binding at a protein-protein interface. For this reason, to predict phosphopeptide selectivity of SH2 domains, Sheinerman *et al.* (2003) created new context-specific substitution matrices based on residues that make important energetic contributions within the SH2 binding site. To be able to systematically predict substrate selectivity of ClpX, we may need to similarly create new matrices that account for the amino acid substitutions we find through our library experiments to be acceptable to retain interactions with ClpX.

Where on the surface of ClpX are its substrate binding sites?

Identifying the binding pockets that accommodate these ClpX-recognition sequences will go hand in hand with the definition of the motifs themselves. Although it is clear that ClpX has multiple substrate binding sites (Siddiqui et al. 2004), it remains largely unknown where on the surface of ClpX these sites lie and which substrate sequences they bind. By examining and perhaps even isolating distinct substrate binding domains, we can probe their specific binding properties. In addition, competition experiments between substrates carrying different motifs will help determine which classes, if any, are binding to the same sites on ClpX.

Are there additional ClpX-adaptor proteins?

Adaptor proteins play an integral role in coordinating changes in environmental conditions with changes in the availability of proteins in the cell (Muffler et al. 1996; Zhou and

Gottesman 1998). From our studies, it appears that a key role of ClpXP is to modify the proteome in response to various stresses. For example, many proteins trapped by ClpXP help cells cope with oxidative stress and shifts between aerobic and anaerobic growth. The majority of these proteins carry primary sequences motifs that target the attached protein for degradation. It is attractive to consider that the accessibility of these sequences to ClpXP may be regulated by an adaptor protein that directly responds to oxygen levels. The same could be true for the stresses of starvation, UV damage, etc. There are many ClpXP substrates, and adaptor proteins are a very powerful mechanism to expand the diversity of sequences that can be recognized by ClpX and control the correct degradation of these proteins.

Does competition between ClpX-adaptor proteins play a regulatory role in the cell?

The identified ClpXP-adaptor proteins all appear to interact with the same site on the N-terminal domain of ClpX. Unidentified adaptor proteins may very well interact with the same site. Competition between these adaptor proteins could play significant regulatory roles in the cell. Up-regulation of one adaptor during certain cellular conditions could enhance delivery of its interacting partners while out-competing other adaptor proteins and thus inhibiting the proteolysis of other substrates. A pioneering experiment to begin to address this question could be to look at the amount of σ^S (an RssB-dependent substrate) captured by $ClpXP^{trap}$ in cells over-expressing SspB. In this case, binding of SspB to the N-domain of ClpX could compete for formation of the RssB-σ^S-ClpX delivery complex. It would be beneficial when analyzing these types of experiments to first examine the expression profiles of the different adaptor proteins. Under what cellular conditions are they up-regulated? What controls this expression?

Are there additional ClpXP substrates under different environmental conditions?

Approximately 50 ClpXP substrates were identified under "normal" growth conditions at 30°C. Many additional ClpXP substrates have been identified under conditions of DNA damage (S. Neher, unpublished data). In addition, the profile of proteins captured by $ClpP^{trap}$ in cells grown at 43°C is quite different than the pattern of those captured at 30°C (unpublished data). How many ClpXP substrates are there? Capturing substrates in $ClpP^{trap}$ while varying environmental conditions such as pH, oxygen availability, nutrient availability, and osmotic pressure will likely identify many more substrates. It is probable that a subset of substrates will be trapped in all the experiments and are proteins that are constitutively degraded.

These various trapping experiments will provide insights into many of the questions posed in this section. For instance, sequence alignment of the new substrates with the previously defined classes of recognition motifs will help delineate the optimal ClpX-binding motifs. In addition, investigating specific substrates that are degraded under one set of conditions but not another may lead to the identification of new adaptor proteins. It is certain that further characterization of these substrates will help us understand the critical role of ClpXP in regulating protein availability in the cell.

Are any of the trapped proteins primarily ClpX-disassembly substrates?

As we discussed in the introduction, ClpX disassembles the hyper-stable MuA-DNA complex. This restructuring activity does not require ClpP, although ClpXP is able to degrade MuA. Recent experiments have shown that during the disassembly of the MuA tetramer, ClpX only contacts certain subunits; it is possible that in vivo these contacted subunits may in fact be degraded (Burton and Baker 2003). It is attractive to imagine that ClpX remodels a number of macromolecular complexes in vivo. For some proteins, it is possible that the critical

function played by ClpX is disassembly, rather than degradation. Substrates that form stable complexes with DNA seems like a reasonable place to begin searching for these targets.

What role does ClpXP play in the regulation of specific substrates?

Characterization of specific ClpXP substrates has expanded our understanding of the important physiological roles of this protease. For example, LexA was caught in the ClpXP^{trap}, and western blots indicated that the autocleavage products but not full-length protein were captured. Further investigation of ClpXP-mediated degradation of LexA cleavage products led to discovery of a role for ClpXP in activation of the SOS response (Neher et al. 2003a). Below are examples of functional groups of trapped substrates; examining the role protein turnover plays in each substrate's regulation will broaden our knowledge of the complex functions of ClpXP.

Proteins with Fe-S centers: Six trapped proteins contain Fe-S centers which can serve as sensors of oxidative stress. For example, the Fe-S cluster of the transcriptional regulator Fnr is oxidized during aerobic growth (Kiley and Beinert 1998), reducing Fnr activity and enhancing its degradation by ClpXP (P. Kiley, unpublished data). Based on these initial studies, ClpXP may degrade proteins whose Fe-S clusters have been damaged by oxidation as a general response to oxidative stress.

Proteins with signal peptides (OmpA): OmpA, an outer membrane porin, was one of the few proteins captured by ClpXP^{trap} that is not normally located in the same cellular compartment as ClpXP. OmpA has an N-terminal recognition signal that overlaps with its signal peptide that permits its export to the periplasm. A large number of signal peptides, including that of OmpC and OmpF, also share this conserved motif, indicating an overlap between ClpX-recognition signals and that of the Sec translocation machinery. It would be very intriguing to investigate whether ClpXP plays a role in protein quality control by degrading proteins with mislocalized secretion tags.

Proteins involved in the stationary phase response: *ClpP*-deleted cells show delayed recovery from stationary phase. A set of ClpXP substrates are proteins that are active during stationary phase. In fact, two of the captured proteins, Crl and DksA, modulate the activity of σ^S (Pratt and Silhavy 1998; Webb et al. 1999). Are these proteins degraded during exponential phase and stabilized in stationary phase in a similar manner as σ^S? Is there an adaptor protein such as RssB that coordinates growth conditions with this temporal degradation? Many questions remain unanswered in the integral role of ClpXP in this environmental stress.

These are just a few examples of intriguing groups of ClpXP substrates. Studying the degradation of each substrate both in vitro and in vivo will likely uncover novel regulatory strategies used by ClpXP to control protein turnover and reveal novel roles for ClpXP in *E. coli* regulatory networks.

Recent advances have provided a wealth of structural and mechanistic information regarding the AAA+ proteases. The next few years will likely lead to precise characterization of binding motifs for all of the protease complexes and many new critical roles these proteases play in regulating the availability of proteins in the cell.

REFERENCES

Ades, S.E. 2004. Control of the alternative sigma factor sigma(E) in *Escherichia coli*. *Curr Opin Microbiol* **7**: 157-62.

Ades, S.E., L.E. Connolly, B.M. Alba, and C.A. Gross. 1999. The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev* **13**: 2449-61.

Alba, B.M., J.A. Leeds, C. Onufryk, C.Z. Lu, and C.A. Gross. 2002. DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev* **16**: 2156-68.

Alba, B.M., H.J. Zhong, J.C. Pelayo, and C.A. Gross. 2001. degS (hhoB) is an essential *Escherichia coli* gene whose indispensable function is to provide sigma (E) activity. *Mol Microbiol* **40**: 1323-33.

Alley, D. 2002. In *Proteolysis in prokaryotes: protein quality control and regulatory principles* (ed. R. Hengge, Bukau, B.), Schwetzingen, Germany.

Almiron, M., A.J. Link, D. Furlong, and R. Kolter. 1992. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* **6**: 2646-54.

Andersen, J.B., C. Sternberg, L.K. Poulsen, S.P. Bjorn, M. Givskov, and S. Molin. 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**: 2240-6.

Anderson, L.A., E. McNairn, T. Lubke, R.N. Pau, D.H. Boxer, and T. Leubke. 2000. ModE-dependent molybdate regulation of the molybdenum cofactor operon moa in *Escherichia coli*. *J Bacteriol* **182**: 7035-43.

Battista, J.R., C.E. Donnelly, T. Ohta, and G.C. Walker. 1990. The SOS response and induced mutagenesis. *Prog Clin Biol Res* **340A**: 169-78.

Bearson, S.M., W.H. Benjamin, Jr., W.E. Swords, and J.W. Foster. 1996. Acid shock induction of RpoS is mediated by the mouse virulence gene mviA of Salmonella typhimurium. *J Bacteriol* **178**: 2572-9.

Becker, G., E. Klauck, and R. Hengge-Aronis. 1999. Regulation of RpoS proteolysis in *Escherichia coli*: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. *Proc Natl Acad Sci U S A* **96**: 6439-44.

Beebe, K.D., J. Shin, J. Peng, C. Chaudhury, J. Khera, and D. Pei. 2000. Substrate recognition through a PDZ domain in tail-specific protease. *Biochemistry* **39**: 3149-55.

Ben-Bassat, A., K. Bauer, S.Y. Chang, K. Myambo, A. Boosman, and S. Chang. 1987. Processing of the initiation methionine from proteins: properties of the

Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol* **169**: 751-7.

Bjellqvist, B., C. Pasquali, F. Ravier, J.C. Sanchez, and D. Hochstrasser. 1993. A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis* **14**: 1357-65.

Bolon, D.N., D.A. Wah, G.L. Hersch, T.A. Baker, and R.T. Sauer. 2004. Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol Cell* **13**: 443-9.

Bouche, S., E. Klauck, D. Fischer, M. Lucassen, K. Jung, and R. Hengge-Aronis. 1998. Regulation of RssB-dependent proteolysis in *Escherichia coli*: a role for acetyl phosphate in a response regulator-controlled process. *Mol Microbiol* **27**: 787-95.

Burton, B.M. and T.A. Baker. 2003. Mu transpososome architecture ensures that unfolding by ClpX or proteolysis by ClpXP remodels but does not destroy the complex. *Chem Biol* **10**: 463-72.

Burton, R.E., S.M. Siddiqui, Y.I. Kim, T.A. Baker, and R.T. Sauer. 2001. Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *Embo J* **20**: 3092-100.

Campbell, E.A., J.L. Tupy, T.M. Gruber, S. Wang, M.M. Sharp, C.A. Gross, and S.A. Darst. 2003. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol Cell* **11**: 1067-78.

Charette, M.F., G.W. Henderson, and A. Markovitz. 1981. ATP hydrolysis-dependent protease activity of the lon (capR) protein of *Escherichia coli* K-12. *Proc Natl Acad Sci U S A* **78**: 4728-32.

Chen, D.Z., D.V. Patel, C.J. Hackbarth, W. Wang, G. Dreyer, D.C. Young, P.S. Margolis, C. Wu, Z.J. Ni, J. Trias, R.J. White, and Z. Yuan. 2000. Actinonin, a naturally occurring antibacterial agent, is a potent deformylase inhibitor. *Biochemistry* **39**: 1256-62.

Chen, P., P. Johnson, T. Sommer, S. Jentsch, and M. Hochstrasser. 1993. Multiple ubiquitin-conjugating enzymes participate in the in vivo degradation of the yeast MAT alpha 2 repressor. *Cell* **74**: 357-69.

Chiba, S., Y. Akiyama, H. Mori, E. Matsuo, and K. Ito. 2000. Length recognition at the N-terminal tail for the initiation of FtsH-mediated proteolysis. *EMBO Rep* **1**: 47-52.

Chin, D.T., S.A. Goff, T. Webster, T. Smith, and A.L. Goldberg. 1988. Sequence of the lon gene in *Escherichia coli*. A heat-shock gene which encodes the ATP-dependent protease La. *J Biol Chem* **263**: 11718-28.

Cowan, C.A. and M. Henkemeyer. 2001. The SH2/SH3 adaptor Grb4 transduces B-ephrin reverse signals. *Nature* **413**: 174-9.

Cowan, C.A. and M. Henkemeyer. 2002. Ephrins in reverse, park and drive. *Trends Cell Biol* **12**: 339-46.

Craigie, R. and K. Mizuuchi. 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* **51**: 493-501.

Damerau, K. and A.C. St John. 1993. Role of Clp protease subunits in degradation of carbon starvation proteins in *Escherichia coli*. *J Bacteriol* **175**: 53-63.

Dartigalongue, C., D. Missiakas, and S. Raina. 2001. Characterization of the *Escherichia coli* sigma E regulon. *J Biol Chem* **276**: 20866-75.

Datsenko, K.A. and B.L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-5.

Davis, B.D., S.M. Luger, and P.C. Tai. 1986. Role of ribosome degradation in the death of starved *Escherichia coli* cells. *J Bacteriol* **166**: 439-45.

De Las Penas, A., L. Connolly, and C.A. Gross. 1997a. SigmaE is an essential sigma factor in *Escherichia coli*. *J Bacteriol* **179**: 6862-4.

De Las Penas, A., L. Connolly, and C.A. Gross. 1997b. The sigmaE-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol Microbiol* **24**: 373-85.

Deuerling, E., H. Patzelt, S. Vorderwulbecke, T. Rauch, G. Kramer, E. Schaffitzel, A. Mogk, A. Schulze-Specking, H. Langen, and B. Bukau. 2003. Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol Microbiol* **47**: 1317-28.

Dixon, h.b.f.a.r.f. 1972. Specific modification of NH2-terminal residues by transamination. *Methods Enzymol* **25**: 409-419.

Domian, I.J., K.C. Quon, and L. Shapiro. 1997. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* **90**: 415-24.

Domian, I.J., A. Reisenauer, and L. Shapiro. 1999. Feedback control of a master bacterial cell-cycle regulator. *Proc Natl Acad Sci U S A* **96**: 6648-53.

Dougan, D.A., A. Mogk, K. Zeth, K. Turgay, and B. Bukau. 2002a. AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett* **529**: 6-10.

Dougan, D.A., B.G. Reid, A.L. Horwich, and B. Bukau. 2002b. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**: 673-83.

Dougan, D.A., E. Weber-Ban, and B. Bukau. 2003. Targeted delivery of an ssrA-tagged substrate by the adaptor protein SspB to its cognate AAA+ protein ClpX. *Mol Cell* **12**: 373-80.

Driessen, A.J., P. Fekkes, and J.P. van der Wolk. 1998. The Sec system. *Curr Opin Microbiol* **1**: 216-22.

Ebel, W., M.M. Skinner, K.P. Dierksen, J.M. Scott, and J.E. Trempy. 1999. A conserved domain in *Escherichia coli* Lon protease is involved in substrate discriminator activity. *J Bacteriol* **181**: 2236-43.

Eisner, G., H.G. Koch, K. Beck, J. Brunner, and M. Muller. 2003. Ligand crowding at a nascent signal sequence. *J Cell Biol* **163**: 35-44.

Engelberg-Kulka, H. and G. Glaser. 1999. Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu Rev Microbiol* **53**: 43-70.

Erickson, J.W. and C.A. Gross. 1989. Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev* **3**: 1462-71.

Fanning, A.S. and J.M. Anderson. 1996. Protein-protein interactions: PDZ domain networks. *Curr Biol* **6**: 1385-8.

Flynn, J.M., I. Levchenko, M. Seidel, S.H. Wickner, R.T. Sauer, and T.A. Baker. 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **98**: 10584-9.

Flynn, J.M., S.B. Neher, Y.I. Kim, R.T. Sauer, and T.A. Baker. 2003. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**: 671-83.

Frank, E.G., D.G. Ennis, M. Gonzalez, A.S. Levine, and R. Woodgate. 1996. Regulation of SOS mutagenesis by proteolysis. *Proc Natl Acad Sci U S A* **93**: 10291-6.

Fu, G.K. and D.M. Markovitz. 1998. The human LON protease binds to mitochondrial promoters in a single-stranded, site-specific, strand-specific manner. *Biochemistry* **37**: 1905-9.

Fu, G.K., M.J. Smith, and D.M. Markovitz. 1997. Bacterial protease Lon is a site-specific DNA-binding protein. *J Biol Chem* **272**: 534-8.

Glickman, M.H. and A. Ciechanover. 2002. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* **82**: 373-428.

Goldberg, A.L. and A.C. St John. 1976. Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annu Rev Biochem* **45**: 747-803.

Gonciarz-Swiatek, M., A. Wawrzynow, S.J. Um, B.A. Learn, R. McMacken, W.L. Kelley, C. Georgopoulos, O. Sliekers, and M. Zylicz. 1999. Recognition, targeting, and hydrolysis of the lambda O replication protein by the ClpP/ClpX protease. *J Biol Chem* **274**: 13999-4005.

Gonzalez, M., F. Rasulova, M.R. Maurizi, and R. Woodgate. 2000. Subunit-specific degradation of the UmuD/D' heterodimer by the ClpXP protease: the role of trans recognition in UmuD' stability. *Embo J* **19**: 5251-8.

Gottesman, S. 1996. Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* **30**: 465-506.

Gottesman, S. 2003. Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **19**: 565-87.

Gottesman, S., W.P. Clark, V. de Crecy-Lagard, and M.R. Maurizi. 1993. ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and in vivo activities. *J Biol Chem* **268**: 22618-26.

Gottesman, S., E. Roche, Y. Zhou, and R.T. Sauer. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**: 1338-47.

Grant, R.A., D.J. Filman, S.E. Finkel, R. Kolter, and J.M. Hogle. 1998. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* **5**: 294-303.

Griaznova, O. and R.R. Traut. 2000. Deletion of C-terminal residues of *Escherichia coli* ribosomal protein L10 causes the loss of binding of one L7/L12 dimer: ribosomes with one L7/L12 dimer are active. *Biochemistry* **39**: 4075-81.

Grimaud, R., M. Kessel, F. Beuron, A.C. Steven, and M.R. Maurizi. 1998. Enzymatic and structural similarities between the *Escherichia coli* ATP- dependent proteases, ClpXP and ClpAP. *J Biol Chem* **273**: 12476-81.

Gross, C.A., C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. 1998. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63**: 141-55.

Guo, F., M.R. Maurizi, L. Esser, and D. Xia. 2002. Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J Biol Chem* **277**: 46743-52.

Harlow, E.a.L., D. 1988. *Antibodies, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Hellman, U., C. Wernstedt, J. Gonez, and C.H. Heldin. 1995. Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal Biochem* **224**: 451-5.

Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839-72.

Hengge-Aronis, R. 1996a. Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* **21**: 887-93.

Hengge-Aronis, R. 1996b. Regulation of Gene Expression During Entry into Stationary Phase. In *Escherichia coli and Salmonella* (ed. F.C. Neidhardt), pp. 1497-1512. American Society for Microbiology, Washington, D.C.

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

Herman, C., S. Prakash, C.Z. Lu, A. Matouschek, and C.A. Gross. 2003. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol Cell* **11**: 659-69.

Herman, C., D. Thevenet, P. Bouloc, G.C. Walker, and R. D'Ari. 1998. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev* **12**: 1348-55.

Herman, C., D. Thevenet, R. D'Ari, and P. Bouloc. 1995. Degradation of sigma 32, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proc Natl Acad Sci U S A* **92**: 3516-20.

Hershko, A. and A. Ciechanover. 1998. The ubiquitin system. *Annu Rev Biochem* **67**: 425-79.

Hochstrasser, M. 1996. Ubiquitin-dependent protein degradation. *Annu Rev Genet* **30**: 405-39.

Hoskins, J.R., S.Y. Kim, and S. Wickner. 2000a. Substrate recognition by the ClpA chaperone component of ClpAP protease. *J Biol Chem* **275**: 35361-7.

Hoskins, J.R., S.K. Singh, M.R. Maurizi, and S. Wickner. 2000b. Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proc Natl Acad Sci U S A* **97**: 8892-7.

Hoskins, J.R., K. Yanagihara, K. Mizuuchi, and S. Wickner. 2002. ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. *Proc Natl Acad Sci U S A* **99**: 11037-42.

Huang, H.C., M.Y. Sherman, O. Kandror, and A.L. Goldberg. 2001. The molecular chaperone DnaJ is required for the degradation of a soluble abnormal protein in *Escherichia coli*. *J Biol Chem* **276**: 3920-8.

Huisman, O., R. D'Ari, and S. Gottesman. 1984. Cell-division control in *Escherichia coli*: specific induction of the SOS function SfiA protein is sufficient to block septation. *Proc Natl Acad Sci U S A* **81**: 4490-4.

Ishihama, A. 1988. Promoter selectivity of prokaryotic RNA polymerases. *Trends Genet* **4**: 282-6.

Ishihama, A. 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**: 499-518.

Ishii, Y. and F. Amano. 2001. Regulation of SulA cleavage by Lon protease by the C-terminal amino acid of SulA, histidine. *Biochem J* **358**: 473-80.

Jenal, U. and T. Fuchs. 1998. An essential protease involved in bacterial cell-cycle control. *Embo J* **17**: 5658-69.

Jishage, M. and A. Ishihama. 1999. Transcriptional organization and in vivo role of the *Escherichia coli* rsd gene, encoding the regulator of RNA polymerase sigma D. *J Bacteriol* **181**: 3768-76.

Johnson, P.R., R. Swanson, L. Rakhilina, and M. Hochstrasser. 1998. Degradation signal masking by heterodimerization of MATalpha2 and MATa1 blocks their mutual destruction by the ubiquitin-proteasome pathway. *Cell* **94**: 217-27.

Jubete, Y., M.R. Maurizi, and S. Gottesman. 1996. Role of the heat shock protein DnaJ in the lon-dependent degradation of naturally unstable proteins. *J Biol Chem* **271**: 30798-803.

Kanehara, K., K. Ito, and Y. Akiyama. 2002. YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev* **16**: 2147-55.

Kang, P.J. and E.A. Craig. 1990. Identification and characterization of a new *Escherichia coli* gene that is a dosage-dependent suppressor of a dnaK deletion mutation. *J Bacteriol* **172**: 2055-64.

Karzai, A.W., E.D. Roche, and R.T. Sauer. 2000. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* **7**: 449-55.

Karzai, A.W., M.M. Susskind, and R.T. Sauer. 1999. SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *Embo J* **18**: 3793-9.

Katayama, Y., S. Gottesman, J. Pumphrey, S. Rudikoff, W.P. Clark, and M.R. Maurizi. 1988. The two-component, ATP-dependent Clp protease of *Escherichia coli*. Purification, cloning, and mutational analysis of the ATP-binding component. *J Biol Chem* **263**: 15226-36.

Keiler, K.C. and R.T. Sauer. 1996. Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J Biol Chem* **271**: 2589-93.

Keiler, K.C., P.R. Waller, and R.T. Sauer. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**: 990-3.

Kenniston, J.A., T.A. Baker, J.M. Fernandez, and R.T. Sauer. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**: 511-20.

Kihara, A., Y. Akiyama, and K. Ito. 1995. FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* **92**: 4532-6.

Kihara, A., Y. Akiyama, and K. Ito. 1997. Host regulation of lysogenic decision in bacteriophage lambda: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proc Natl Acad Sci U S A* **94**: 5544-9.

Kiley, P.J. and H. Beinert. 1998. Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster. *FEMS Microbiol Rev* **22**: 341-52.

Kim, D.Y. and K.K. Kim. 2003. Crystal structure of ClpX molecular chaperone from Helicobacter pylori. *J Biol Chem* **278**: 50664-70.

Kim, S.J., Y.H. Han, I.H. Kim, and H.K. Kim. 1999. Involvement of ArcA and Fnr in expression of *Escherichia coli* thiol peroxidase gene. *IUBMB Life* **48**: 215-8.

Kim, Y.I., R.E. Burton, B.M. Burton, R.T. Sauer, and T.A. Baker. 2000. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**: 639-48.

Konieczny, I. and D.R. Helinski. 1997. The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone. *Proc Natl Acad Sci U S A* **94**: 14378-82.

Krieg, U.C., P. Walter, and A.E. Johnson. 1986. Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. *Proc Natl Acad Sci U S A* **83**: 8604-8.

Kullander, K. and R. Klein. 2002. Mechanisms and functions of Eph and ephrin signalling. *Nat Rev Mol Cell Biol* **3**: 475-86.

Kuriyan, J. and D. Cowburn. 1997. Modular peptide recognition domains in eukaryotic signaling. *Annu Rev Biophys Biomol Struct* **26**: 259-88.

Kwon, Y.T., Y. Reiss, V.A. Fried, A. Hershko, J.K. Yoon, D.K. Gonda, P. Sangan, N.G. Copeland, N.A. Jenkins, and A. Varshavsky. 1998. The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc Natl Acad Sci U S A* **95**: 7898-903.

Laachouch, J.E., L. Desmet, V. Geuskens, R. Grimaud, and A. Toussaint. 1996. Bacteriophage Mu repressor as a target for the *Escherichia coli* ATP- dependent Clp Protease. *Embo J* **15**: 437-44.

Lee, Y.Y., C.F. Chang, C.L. Kuo, M.C. Chen, C.H. Yu, P.I. Lin, and W.F. Wu. 2003. Subunit oligomerization and substrate recognition of the *Escherichia coli* ClpYQ (HslUV) protease implicated by in vivo protein-protein interactions in the yeast two-hybrid system. *J Bacteriol* **185**: 2393-401.

Lehnherr, H. and M.B. Yarmolinsky. 1995. Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. *Proc Natl Acad Sci U S A* **92**: 3274-7.

Levchenko, I., R.A. Grant, D.A. Wah, R.T. Sauer, and T.A. Baker. 2003. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol Cell* **12**: 365-72.

Levchenko, I., L. Luo, and T.A. Baker. 1995. Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* **9**: 2399-408.

Levchenko, I., M. Seidel, R.T. Sauer, and T.A. Baker. 2000. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**: 2354-6.

Levchenko, I., C.K. Smith, N.P. Walsh, R.T. Sauer, and T.A. Baker. 1997a. PDZ-like domains mediate binding specificity in the Clp/Hsp100 family of chaperones and protease regulatory subunits. *Cell* **91**: 939-47.

Levchenko, I., M. Yamauchi, and T.A. Baker. 1997b. ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes Dev* **11**: 1561-72.

Lipinska, B., A. Podhajska, and K. Taylor. 1980. Synthesis and decay of lambda DNA replication proteins in minicells. *Biochem Biophys Res Commun* **92**: 120-6.

Little, J.W. 1983. Variations in the in vivo stability of LexA repressor during the SOS regulatory cycle. In *Cellular Responses to DNA Damage*.

Little, J.W., S.H. Edmiston, L.Z. Pacelli, and D.W. Mount. 1980. Cleavage of the *Escherichia coli* lexA protein by the recA protease. *Proc Natl Acad Sci U S A* **77**: 3225-9.

Lu, Q., E.E. Sun, R.S. Klein, and J.G. Flanagan. 2001. Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. *Cell* **105**: 69-79.

Luo, Y., R.A. Pfuetzner, S. Mosimann, M. Paetzel, E.A. Frey, M. Cherney, B. Kim, J.W. Little, and N.C. Strynadka. 2001. Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **106**: 585-94.

Lupas, A., J.M. Flanagan, T. Tamura, and W. Baumeister. 1997. Self-compartmentalizing proteases. *Trends Biochem Sci* **22**: 399-404.

Lupas, A.N. and J. Martin. 2002. AAA proteins. *Curr Opin Struct Biol* **12**: 746-53.

Lynch, A.S. 1996. Responses to Molecular Oxygen. In *Escherichia coli and Salmonella* (ed. F.C. Neidhardt), pp. 1526-1538. American Society for Microbiology, Washington, D.C.

Makovets, S., A.J. Titheradge, and N.E. Murray. 1998. ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol Microbiol* **28**: 25-35.

Martinez, A. and R. Kolter. 1997. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol* **179**: 5188-94.

Maurizi, M.R., W.P. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers, and S. Gottesman. 1990a. Sequence and structure of Clp P, the proteolytic component of the ATP- dependent Clp protease of *Escherichia coli*. *J Biol Chem* **265**: 12536-45.

Maurizi, M.R., W.P. Clark, S.H. Kim, and S. Gottesman. 1990b. Clp P represents a unique family of serine proteases. *J Biol Chem* **265**: 12546-52.

Maurizi, M.R. and F. Rasulova. 2002. Degradation of L-glutamate dehydrogenase from *Escherichia coli*: allosteric regulation of enzyme stability. *Arch Biochem Biophys* **397**: 206-16.

Maurizi, M.R., M.W. Thompson, S.K. Singh, and S.H. Kim. 1994. Endopeptidase Clp: ATP-dependent Clp protease from *Escherichia coli*. *Methods Enzymol* **244**: 314-31.

Mecsas, J., P.E. Rouviere, J.W. Erickson, T.J. Donohue, and C.A. Gross. 1993. The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev* **7**: 2618-28.

Milla, M.E., B.M. Brown, and R.T. Sauer. 1993. P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences. *Protein Sci* **2**: 2198-205.

Miller, J.H. 1972. *Experiments in Molecular Genetics*. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Missiakas, D., M.P. Mayer, M. Lemaire, C. Georgopoulos, and S. Raina. 1997. Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol Microbiol* **24**: 355-71.

Missiakas, D., F. Schwager, J.M. Betton, C. Georgopoulos, and S. Raina. 1996. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *Embo J* **15**: 6899-909.

Mizusawa, S. and S. Gottesman. 1983. Protein degradation in *Escherichia coli*: the lon gene controls the stability of sulA protein. *Proc Natl Acad Sci U S A* **80**: 358-62.

Msadek, T., V. Dartois, F. Kunst, M.L. Herbaud, F. Denizot, and G. Rapoport. 1998. ClpP of Bacillus subtilis is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. *Mol Microbiol* **27**: 899-914.

Muffler, A., D. Fischer, S. Altuvia, G. Storz, and R. Hengge-Aronis. 1996. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in *Escherichia coli*. *Embo J* **15**: 1333-9.

Muller, M., H.G. Koch, K. Beck, and U. Schafer. 2001. Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane. *Prog Nucleic Acid Res Mol Biol* **66**: 107-57.

Nagai, H., H. Yuzawa, M. Kanemori, and T. Yura. 1994. A distinct segment of the sigma 32 polypeptide is involved in DnaK-mediated negative control of the heat shock response in *Escherichia coli*. *Proc Natl Acad Sci U S A* **91**: 10280-4.

Neher, S.B., J.M. Flynn, R.T. Sauer, and T.A. Baker. 2003a. Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes Dev* **17**: 1084-9.

Neher, S.B., R.T. Sauer, and T.A. Baker. 2003b. Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease. *Proc Natl Acad Sci U S A* **100**: 13219-24.

Neuwald, A.F., L. Aravind, J.L. Spouge, and E.V. Koonin. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**: 27-43.

Ogura, T. and A.J. Wilkinson. 2001. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**: 575-97.

Okusu, H., D. Ma, and H. Nikaido. 1996. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol* **178**: 306-8.

Ortega, J., S.K. Singh, T. Ishikawa, M.R. Maurizi, and A.C. Steven. 2000. Visualization of substrate binding and translocation by the ATP- dependent protease, ClpXP. *Mol Cell* **6**: 1515-21.

Parsell, D.A. and R.T. Sauer. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J Biol Chem* **264**: 7590-5.

Parsell, D.A., K.R. Silber, and R.T. Sauer. 1990. Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev* **4**: 277-86.

Patzelt, H., S. Rudiger, D. Brehmer, G. Kramer, S. Vorderwulbecke, E. Schaffitzel, A. Waitz, T. Hesterkamp, L. Dong, J. Schneider-Mergener, B. Bukau, and E. Deuerling. 2001. Binding specificity of *Escherichia coli* trigger factor. *Proc Natl Acad Sci U S A* **98**: 14244-9.

Pelham, H.R. 1986. Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46**: 959-61.

Pellicer, M.T., C. Fernandez, J. Badia, J. Aguilar, E.C. Lin, and L. Baldom. 1999a. Cross-induction of glc and ace operons of *Escherichia coli* attributable to

pathway intersection. Characterization of the glc promoter. *J Biol Chem* **274**: 1745-52.

Pellicer, M.T., A.S. Lynch, P. De Wulf, D. Boyd, J. Aguilar, and E.C. Lin. 1999b. A mutational study of the ArcA-P binding sequences in the aldA promoter of *Escherichia coli*. *Mol Gen Genet* **261**: 170-6.

Persuh, M., K. Turgay, I. Mandic-Mulec, and D. Dubnau. 1999. The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch. *Mol Microbiol* **33**: 886-94.

Petersen, C. 1990. *Escherichia coli* ribosomal protein L10 is rapidly degraded when synthesized in excess of ribosomal protein L7/L12. *J Bacteriol* **172**: 431-6.

Porankiewicz, J., J. Wang, and A.K. Clarke. 1999. New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. *Mol Microbiol* **32**: 449-58.

Pratt, L.A. and T.J. Silhavy. 1996. The response regulator SprE controls the stability of RpoS. *Proc Natl Acad Sci U S A* **93**: 2488-92.

Pratt, L.A. and T.J. Silhavy. 1998. Crl stimulates RpoS activity during stationary phase. *Mol Microbiol* **29**: 1225-36.

Quon, K.C., G.T. Marczynski, and L. Shapiro. 1996. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**: 83-93.

Raina, S., D. Missiakas, and C. Georgopoulos. 1995. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of *Escherichia coli*. *Embo J* **14**: 1043-55.

Rao, H., F. Uhlmann, K. Nasmyth, and A. Varshavsky. 2001. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410**: 955-9.

Reiss, Y., D. Kaim, and A. Hershko. 1988. Specificity of binding of NH2-terminal residue of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J Biol Chem* **263**: 2693-8.

Rezuchova, B., H. Miticka, D. Homerova, M. Roberts, and J. Kormanec. 2003. New members of the *Escherichia coli* sigmaE regulon identified by a two-plasmid system. *FEMS Microbiol Lett* **225**: 1-7.

Roberts, J.W. and C.W. Roberts. 1975. Proteolytic cleavage of bacteriophage lambda repressor in induction. *Proc Natl Acad Sci U S A* **72**: 147-51.

Robinson, C.R. and R.T. Sauer. 1996. Equilibrium stability and sub-millisecond refolding of a designed single-chain Arc repressor. *Biochemistry* **35**: 13878-84.

Roche, E.D. and R.T. Sauer. 1999. SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *Embo J* **18**: 4579-89.

Rohrwild, M., O. Coux, H.C. Huang, R.P. Moerschell, S.J. Yoo, J.H. Seol, C.H. Chung, and A.L. Goldberg. 1996. HslV-HslU: A novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc Natl Acad Sci U S A* **93**: 5808-13.

Rosenfeld, J., J. Capdevielle, J.C. Guillemot, and P. Ferrara. 1992. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem* **203**: 173-9.

Roudiak, S.G. and T.E. Shrader. 1998. Functional role of the N-terminal region of the Lon protease from Mycobacterium smegmatis. *Biochemistry* **37**: 11255-63.

Rouviere, P.E., A. De Las Penas, J. Mecsas, C.Z. Lu, K.E. Rudd, and C.A. Gross. 1995. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *Embo J* **14**: 1032-42.

Rudiger, S., L. Germeroth, J. Schneider-Mergener, and B. Bukau. 1997. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *Embo J* **16**: 1501-7.

Ryan, K.R., S. Huntwork, and L. Shapiro. 2004. Recruitment of a cytoplasmic response regulator to the cell pole is linked to its cell cycle-regulated proteolysis. *Proc Natl Acad Sci U S A* **101**: 7415-20.

Ryan, K.R., E.M. Judd, and L. Shapiro. 2002. The CtrA response regulator essential for Caulobacter crescentus cell-cycle progression requires a bipartite degradation signal for temporally controlled proteolysis. *J Mol Biol* **324**: 443-55.

Schirmer, E.C., J.R. Glover, M.A. Singer, and S. Lindquist. 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci* **21**: 289-96.

Schlothauer, T., A. Mogk, D.A. Dougan, B. Bukau, and K. Turgay. 2003. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc Natl Acad Sci U S A* **100**: 2306-11.

Schmucker, D. and S.L. Zipursky. 2001. Signaling downstream of Eph receptors and ephrin ligands. *Cell* **105**: 701-4.

Schweder, T., K.H. Lee, O. Lomovskaya, and A. Matin. 1996. Regulation of *Escherichia coli* starvation sigma factor (sigma s) by ClpXP protease. *J Bacteriol* **178**: 470-6.

Sheinerman, F.B., B. Al-Lazikani, and B. Honig. 2003. Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J Mol Biol* **334**: 823-41.

Shinagawa, H., H. Iwasaki, T. Kato, and A. Nakata. 1988. RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci U S A* **85**: 1806-10.

Siddiqui, S.M. 2004. Dissecting the Steps of Substrate Processing by the Energy-Dependent Protease ClpXP. In *Biology*. MIT, Cambridge.

Siddiqui, S.M., R.T. Sauer, and T.A. Baker. 2004. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev* **18**: 369-74.

Singh, S.K., R. Grimaud, J.R. Hoskins, S. Wickner, and M.R. Maurizi. 2000. Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP. *Proc Natl Acad Sci U S A* **97**: 8898-903.

Singh, S.K., J. Rozycki, J. Ortega, T. Ishikawa, J. Lo, A.C. Steven, and M.R. Maurizi. 2001. Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *J Biol Chem* **276**: 29420-9.

Skerker, J.M. and M.T. Laub. 2004. Cell-cycle progression and the generation of asymmetry in Caulobacter crescentus. *Nat Rev Microbiol* **2**: 325-37.

Songyang, Z., A.S. Fanning, C. Fu, J. Xu, S.M. Marfatia, A.H. Chishti, A. Crompton, A.C. Chan, J.M. Anderson, and L.C. Cantley. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **275**: 73-7.

Songyang, Z., S.E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W.G. Haser, F. King, T. Roberts, S. Ratnofsky, R.J. Lechleider, and et al. 1993. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**: 767-78.

Songyang, Z., S.E. Shoelson, J. McGlade, P. Olivier, T. Pawson, X.R. Bustelo, M. Barbacid, H. Sabe, H. Hanafusa, T. Yi, and et al. 1994. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol Cell Biol* **14**: 2777-85.

Straus, D., W. Walter, and C.A. Gross. 1990. DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes Dev* **4**: 2202-9.

Studemann, A., M. Noirclerc-Savoye, E. Klauck, G. Becker, D. Schneider, and R. Hengge. 2003. Sequential recognition of two distinct sites in sigma(S) by the proteolytic targeting factor RssB and ClpX. *Embo J* **22**: 4111-20.

Su, Z., P. Xu, and F. Ni. 2004. Single phosphorylation of Tyr304 in the cytoplasmic tail of ephrin B2 confers high-affinity and bifunctional binding to both the SH2 domain of Grb4 and the PDZ domain of the PDZ-RGS3 protein. *Eur J Biochem* **271**: 1725-36.

Surette, M.G., S.J. Buch, and G. Chaconas. 1987. Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* **49**: 253-62.

Sutton, M.D., B.T. Smith, V.G. Godoy, and G.C. Walker. 2000. The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu Rev Genet* **34**: 479-497.

Thompson, M.W. and M.R. Maurizi. 1994. Activity and specificity of *Escherichia coli* ClpAP protease in cleaving model peptide substrates. *J Biol Chem* **269**: 18201-8.

Tilly, K., J. Spence, and C. Georgopoulos. 1989. Modulation of stability of the *Escherichia coli* heat shock regulatory factor sigma. *J Bacteriol* **171**: 1585-9.

Tobias, J.W., T.E. Shrader, G. Rocap, and A. Varshavsky. 1991. The N-end rule in bacteria. *Science* **254**: 1374-7.

Tomoyasu, T., F. Arsene, T. Ogura, and B. Bukau. 2001. The C terminus of sigma(32) is not essential for degradation by FtsH. *J Bacteriol* **183**: 5911-7.

Tomoyasu, T., J. Gamer, B. Bukau, M. Kanemori, H. Mori, A.J. Rutman, A.B. Oppenheim, T. Yura, K. Yamanaka, H. Niki, and et al. 1995. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *Embo J* **14**: 2551-60.

Tomoyasu, T., T. Ogura, T. Tatsuta, and B. Bukau. 1998. Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. *Mol Microbiol* **30**: 567-81.

Tu, G.F., G.E. Reid, J.G. Zhang, R.L. Moritz, and R.J. Simpson. 1995. C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *J Biol Chem* **270**: 9322-6.

Turgay, K., L.W. Hamoen, G. Venema, and D. Dubnau. 1997. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of Bacillus subtilis. *Genes Dev* **11**: 119-28.

Uchiumi, T., K. Hori, T. Nomura, and A. Hachimori. 1999. Replacement of L7/L12.L10 protein complex in *Escherichia coli* ribosomes with the eukaryotic counterpart changes the specificity of elongation factor binding. *J Biol Chem* **274**: 27578-82.

Uhlmann, F., D. Wernic, M.A. Poupart, E.V. Koonin, and K. Nasmyth. 2000. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103**: 375-86.

Varshavsky, A. 1992. The N-end rule. *Cell* **69**: 725-35.

Varshavsky, A. 1996. The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci U S A* **93**: 12142-9.

Voges, D., P. Zwickl, and W. Baumeister. 1999. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* **68**: 1015-68.

Wah, D.A., I. Levchenko, T.A. Baker, and R.T. Sauer. 2002. Characterization of a specificity factor for an AAA+ ATPase: assembly of SspB dimers with ssrA-tagged proteins and the ClpX hexamer. *Chem Biol* **9**: 1237-45.

Wah, D.A., I. Levchenko, G.E. Rieckhof, D.N. Bolon, T.A. Baker, and R.T. Sauer. 2003. Flexible linkers leash the substrate binding domain of SspB to a peptide module that stabilizes delivery complexes with the AAA+ ClpXP protease. *Mol Cell* **12**: 355-63.

Walsh, N.P., B.M. Alba, B. Bose, C.A. Gross, and R.T. Sauer. 2003. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113**: 61-71.

Wang, J., J.A. Hartling, and J.M. Flanagan. 1997. The structure of ClpP at 2.3 A resolution suggests a model for ATP- dependent proteolysis. *Cell* **91**: 447-56.

Wang, J., J.J. Song, M.C. Franklin, S. Kamtekar, Y.J. Im, S.H. Rho, I.S. Seong, C.S. Lee, C.H. Chung, and S.H. Eom. 2001. Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure (Camb)* **9**: 177-84.

Wang, L., M. Elliott, and T. Elliott. 1999. Conditional stability of the HemA protein (glutamyl-tRNA reductase) regulates heme biosynthesis in Salmonella typhimurium. *J Bacteriol* **181**: 1211-9.

Wawrzynow, A., D. Wojtkowiak, J. Marszalek, B. Banecki, M. Jonsen, B. Graves, C. Georgopoulos, and M. Zylicz. 1995. The ClpX heat-shock protein of *Escherichia coli*, the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. *Embo J* **14**: 1867-77.

Webb, C., M. Moreno, M. Wilmes-Riesenberg, R. Curtiss, 3rd, and J.W. Foster. 1999. Effects of DksA and ClpP protease on sigma S production and virulence in Salmonella typhimurium. *Mol Microbiol* **34**: 112-23.

Weber-Ban, E.U., B.G. Reid, A.D. Miranker, and A.L. Horwich. 1999. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**: 90-3.

Wegrzyn, G., A. Pawlowicz, and K. Taylor. 1992. Stability of coliphage lambda DNA replication initiator, the lambda O protein. *J Mol Biol* **226**: 675-80.

White, D.G., J.D. Goldman, B. Demple, and S.B. Levy. 1997. Role of the acrAB locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in *Escherichia coli*. *J Bacteriol* **179**: 6122-6.

Wickner, S., S. Gottesman, D. Skowyra, J. Hoskins, K. McKenney, and M.R. Maurizi. 1994. A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc Natl Acad Sci U S A* **91**: 12218-22.

Wiegert, T. and W. Schumann. 2001. SsrA-mediated tagging in Bacillus subtilis. *J Bacteriol* **183**: 3885-9.

Withey, J.H. and D.I. Friedman. 2003. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu Rev Microbiol* **57**: 101-23.

Wojtkowiak, D., C. Georgopoulos, and M. Zylicz. 1993. Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of *Escherichia coli*. *J Biol Chem* **268**: 22609-17.

Wojtyra, U.A., G. Thibault, A. Tuite, and W.A. Houry. 2003. The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J Biol Chem* **278**: 48981-90.

Wu, W.F., Y. Zhou, and S. Gottesman. 1999. Redundant in vivo proteolytic activities of *Escherichia coli* Lon and the ClpYQ (HslUV) protease. *J Bacteriol* **181**: 3681-7.

Wyatt, W.M. and H. Inokuchi. 1974. Stability of lambda O and P replication functions. *Virology* **58**: 313-5.

Yaffe, M.B., G.G. Leparc, J. Lai, T. Obata, S. Volinia, and L.C. Cantley. 2001. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* **19**: 348-53.

Yakhnin, A.V., L.M. Vinokurov, A.K. Surin, and Y.B. Alakhov. 1998. Green fluorescent protein purification by organic extraction. *Protein Expr Purif* **14**: 382-6.

Yura, T., Kanemori, M., Morita, M.T. 2000. The heat shock response: regulation and function. In *Bacterial Stress Responses* (ed. G. Storz, Hengge-Aronis, R.). ASM Press, Washington, D.C.

Zhou, Y. and S. Gottesman. 1998. Regulation of proteolysis of the stationary-phase sigma factor RpoS. *J Bacteriol* **180**: 1154-8.

Zhou, Y., S. Gottesman, J.R. Hoskins, M.R. Maurizi, and S. Wickner. 2001. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. *Genes Dev* **15**: 627-37.

Zylicz, M., K. Liberek, A. Wawrzynow, and C. Georgopoulos. 1998. Formation of the preprimosome protects lambda O from RNA transcription-dependent proteolysis by ClpP/ClpX. *Proc Natl Acad Sci U S A* **95**: 15259-63.