Kevin H. Wang

B.A., Molecular & Cellular Biology
University of California, Berkeley 2003

Submitted to the department of biology in partial
fulfillment of the requirements for the degree of

Doctor of Philosophy
At the
Massachusetts Institute of Technology

September 2008

Signature of Author:

Kevin H. Wang
Dept. of Biology
August 12, 2008

Certified by:

Tania A. Baker
Professor of Biology
Thesis supervisor

Accepted by:

Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

Tuning Strength and Specificity in the N-end Rule

by Kevin H. Wang

Submitted to the Department of Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology

Abstract

Protein degradation is a central mechanism in the regulation of gene expression and activity. Proteolysis regulates not only homeostatic activities, but also the cell's responses to stress. A recurring question underlying this regulatory process is the specificity of substrate selection by the proteolytic machinery.

I designed an unbiased selection to isolate N-terminal degradation sequences *in vivo*, which led to a collection of N-end rule signals. The N-end rule describes how the identity of a protein's N-terminal residue determines its metabolic stability. In *E. coli*, ClpAP is the principal protease that degrades proteins bearing an N-terminal phenylalanine, tyrosine, tryptophan, or leucine reisdue. The ClpS adaptor, which displays homology to eukaryotic ubiquitin ligases that recognize N-end signals, is a recently discovered component of the bacterial N-end rule. Using the collection of N-end signals, I was able to demonstrate that ClpS enhances N-end degradation by ClpAP but is not required *in vivo* or *in vitro*.

The collection of N-end signals also provided insight into the role of sequence context in the N-end rule. Specifically, acidic residues and the length of the N-end signal affect degradation rates *in vitro*. These defective N-end signals also allowed us to separately define recognition specificities of ClpS and ClpAP. Whereas ClpS bound poorly to acidic N-end signals, ClpAP was unable to degrade substrates with short N-end sequences.

Although two decades of biochemical and cellular data support the importance of the N-terminal residue in N-end degradation, there has been no structural information explaining how a single residue is recognized as a degradation signal. To this end, we solved a cocrystal structure of ClpS in complex with an N-end peptide. ClpS uses an extensive hydrogen bonding network to dock the α-amino group and a cavity lined with hydrophobic residues to recognize the N-terminal residue. Furthermore, mutation of the hydrophobic cavity altered the specificity of ClpS toward N-terminal residues.

Together these findings attribute molecular functions to ClpS and ClpAP in the bacterial N-end rule and define sequence rules for the N-end signal. Furthermore, this work provides the tools and background for investigating the mechanism of substrate delivery by ClpS to ClpAP.

Thesis supervisor: Tania A. Baker

Acknowledgements

In no particular order, I wish to thank the following:

Members of the Baker lab, past and present, for their desire to maintain an intellectually stimulating, supportive, and entertaining research environment.

My advisor **Tania Baker**, an awesome experimental scientist and caring advisor who is always thinking of how to bring out the most potential in her students. If the urban legend is indeed true that graduate students adopt habits of their advisors, then I have much to look forward to.

My co-advisor Bob "the Crusher" Sauer, whose penchant for designing projects and constant motivation (especially when it comes to structure building) inspired me to go further than I thought possible.

Frank Solomon and Steve Bell, members of my thesis committee who remind me periodically to take a step back and observe the bigger picture of my work.

The MIT biology program, which actually can be a very cohesive and nurturing environment, contrary to popular belief. Special thanks to biograd 2003, a class of capable and intelligent people who understand comraderie and support.

Members of the Sauer lab for scientific input, and the rest of the 5th floor for sustaining an open and entertaining research environment.

My parents and brother, who remind me that home is always in San Diego.

Brian Sato and Mike Heinrichs, friends from high school who will be friends for a lifetime.

Eric Chun, for compulsive adventures and reminders of nostalgia.

TABLE OF CONTENTS

List of Figures

Chapter One

Introduction to substrate selection by energy-dependent proteases

Overview of cellular roles for protein degradation

Organisms utilize a diverse array of biological activities to thrive in their environments and adapt to changing conditions. One facet of these activities is carried out by proteins that function as structural components, catalysts, or sensors for a particular response. A consequence of this constant optimization is the need to regulate the timing and magnitude of responses to stress; for example, biological activities that promote recovery from DNA damage are debilitating for growth in nutrient-rich conditions (Sutton et al. 2000). To regulate protein function, cells have developed methods of controlling protein synthesis, degradation, and activity.

Protein degradation is a process that is antagonized by several properties of folded proteins. Translation is an energy-consuming procedure that results in peptide bonds that are thermodynamically resistant to chemical degradation (Daniel et al. 1996). This hurdle is accompanied by the problem of accessibility, because a significant portion of residues are buried upon protein folding, reducing the number of labile sites available to agents capable of cleaving peptide linkages. Complete protein degradation therefore requires unfolding followed by peptide bond hydrolysis, a role well-suited to energy-dependent proteases.

Several well-characterized examples highlight the pivotal role of proteolysis during cellular stress and growth. The E. coli transcription factor σ^{32} is constantly degraded until the cells encounter heat shock, which stabilizes σ^{32} and allows σ^{32} to activate expression of genes encoding chaperones to aid the bacteria in resisting damage from the shock. The wasteful expenditure of energy required to constantly synthesize and degrade σ^{32} implies that this response is more beneficial to bacteria

than de novo synthesis of σ^{32} upon heat shock. The conclusion from this observation is that the rate of degradation *in vivo* can be tightly controlled: it is rapid enough to prevent σ^{32} from activating its transcriptional targets but can be suppressed to allow for σ^{32} to accumulate in the cell.

Progression through the eukaryotic cell cycle requires exquisite temporal organization. Each phase of the cell cycle is driven by signaling pathways that use cyclin proteins to define and choreograph the activities that must be carried out. In order to progress from one phase to another, cyclins from a previous phase must be first removed by degradation. Additionally, without complete turnover of proteins that link chromosomes during metaphase, the fidelity of chromosome partitioning deteriorates and leads to aneuploidy and genomic instability.

Proteolysis serves to eliminate damaged or unfolded proteins and replenish pools of amino acids in the cell (Visick and Clarke 1995). Because direct detection of unfolded protein degradation *in vivo* is technically difficult, examples of protein turnover can be inferred from studies of bulk protein turnover in cells grown at temperatures where a significant amount of cellular proteins may be denatured (Parag et al. 1987). An elegant study showed that the mitochondrial orthologue of Lon protease preferentially degraded aconitase enzyme when aconitase was first oxidized by hydroxyl radicals (Bota and Davies 2002). Another study of a missorted mitochondrial P450 enzyme demonstrated the importance of mitochondrial Lon protease in preventing aggregation (Savel'ev et al. 1998).

From the examples above, it is clear that energy-dependent proteases recognize a broad array of substrates. How do these enzymes select their targets? What is the

molecular signature that targets a protein for destruction? In this introduction, I will discuss the strategies used by energy-dependent proteases to carry out protein unfolding and degradation. I will also describe the regulation of substrate selection by the ClpAP protease and ClpS adaptor in the N-end rule degradation pathway.

Architecture of energy-dependent proteases

The 20S proteasome from eukaryotic cells and archaea share a common assemblage with energy-dependent proteases from bacteria, hinting that form begets function. These molecular machines are composed of an ATP-dependent unfolding enzyme coupled to a proteolytic chamber. *E. coli* contains five such proteases: ClpAP, ClpXP, HslUV, Lon, and FtsH (Gottesman 1996). Lon and FtsH are homo-hexameric complexes composed of subunits that contain the ATPase and proteolytic chamber domains on a single polypeptide chain (Chin et al. 1988; Tomoyasu et al. 1995). In contrast, ClpA, ClpX, and HslU form hexameric ATP-dependent unfoldases that complex with their cognate proteolytic chambers ClpP or HslV (Katayama et al. 1988; Gottesman et al. 1993; Rohrwild et al. 1996) (Fig. 1.1A).

ClpA is a member of the Clp/Hsp subclass of enzymes that exists as part of the AAA+ (ATPases Associated with a variety of Activities) family of ATPases (Fig. 1.1B). The Clp/Hsp ATPases carry out processes involved in thermotolerance and protein quality control (Parsell and Lindquist 1993). ClpA contains two ATP-binding domains although only one hydrolyzes ATP (Singh and Maurizi 1994; Seol et al. 1995). In both domains, residues required for ATP binding are contributed by neighboring subunits such that ATP binding promotes oligomerization (Guo et al. 2002b). The N-terminal

domain in ClpA is attached to one ATP-binding domain by a flexible linker and binds substrates and the adaptor protein ClpS (Lo et al. 2001; Guo et al. 2002a; Zeth et al. 2002; Hinnerwisch et al. 2005b).

Protein unfolding by ClpA is thought to result from the mechanical strain exerted on the substrate as ClpA attempts to pull it through the narrow orifice in the ClpA hexamer. This orifice is concentric with the axial entry gate of the ClpP chamber, which consists of two ClpP heptameric caps (Fig. 1.1C). The ClpP gate is ~10 Å wide, preventing folded proteins from diffusing into the chamber (Wang et al. 1997). Substrate denaturation by ClpA is therefore required to translocate a polypeptide chain in ClpP. In addition, denaturation exposes peptide bonds normally buried in the hydrophobic core of the substrate in solution. These properties of substrate processing imply that ClpA carries out substrate selection whereas ClpP acts as the catalytic protease that depends on ClpA for access to unfolded polypeptides.

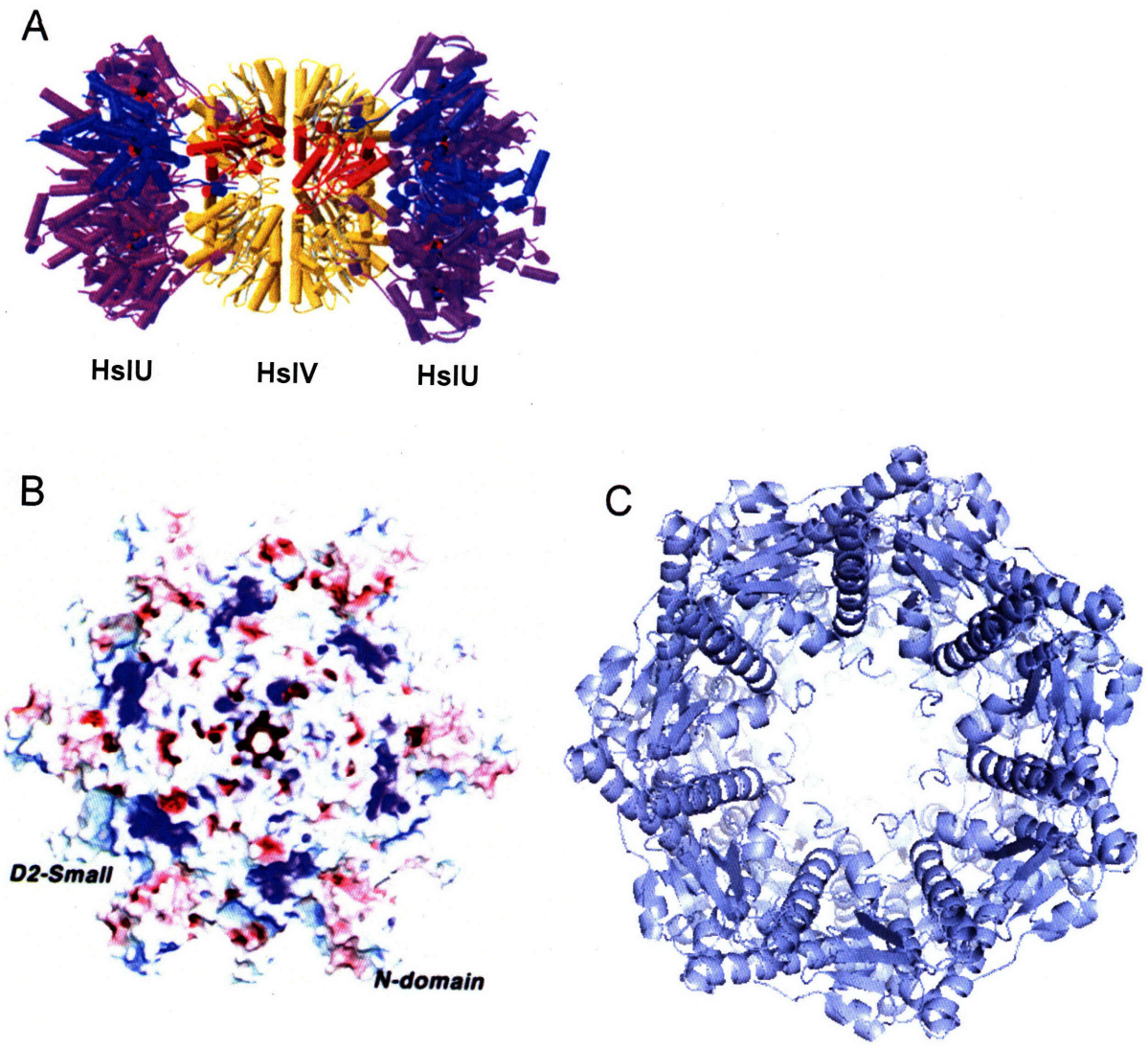


Figure 1.1. Architecture of Clp/Hsp100 proteases. (A) HslU hexamers dock on opposite ends of HslV (Sousa et al. 2000). (B) Top view of the ClpA hexamer ($ClpA_6$) modeled from the crystal structure of the monomer. The N-domain is the N-terminal region that leads into the first ATP-binding domain, and D2-small is the second ATP-binding domain (Guo et al. 2002a). (C) Axial view of $ClpP_{14}$; this view would dock to the bottom of $ClpA_6$ as shown in A (Wang et al. 1997).

Recognition mechanisms for proteolysis

The eukaryotic ubiquitin tagging system

What signals specifically target substrates to a particular protease? Eukaryotic degradation signals are diverse, including but not limited to phosphorylated residues (Feldman et al. 1997; Skowyra et al. 1997; Nash et al. 2001), sugar groups (Yoshida et al. 2002; Yoshida et al. 2003), and hydroxylated proline residues (Bruick and McKnight 2001; Ivan et al. 2001). These signals are recognized by ubiquitin ligases (E3 enzymes) via specific contacts with posttranslational modifications and residue side chains. Structural studies have subsequently provided valuable insight into the diversity of recognition by E3 enzymes, such as toward phosphorylated signals for cyclin E and the DNA replication inhibitor Sic1 (Orlicky et al. 2003; Hao et al. 2007), chitobiose sugar binding (Mizushima et al. 2004), and hydroxylated proline in HIF-1 (hypoxia-inducible factor) (Hon et al. 2002; Min et al. 2002) (Fig. 1.2).

The number of E3 enzymes is constantly increasing as more are discovered since each E3 is specific for a class of degradation signals; as of 2003 there were more than 60 E3 proteins in the yeast proteome (Crews 2003). E3 enzymes recruit substrates to ubiquitin-conjugating enzymes (E2), which are in turn recipients of activated ubiquitin monomers from the E1 ubiquitin-activating enzyme. E1 enzymes transfer ubiquitin monomers to E2 enzymes by thioester transfer (Scheffner et al. 1995). The end result of this process is the covalent attachment of ubiquitin monomers to lysine residues on substrates (Passmore and Barford 2004) (Fig. 1.3). In contrast to the diversity of E3 proteins, there is only one E1 enzyme in eukaryotes called Uba1 and 10 to 20 E2

enzymes (Pickart 2001), implying that regulation of protein degradation occurs at the level of E3 substrate specificity.

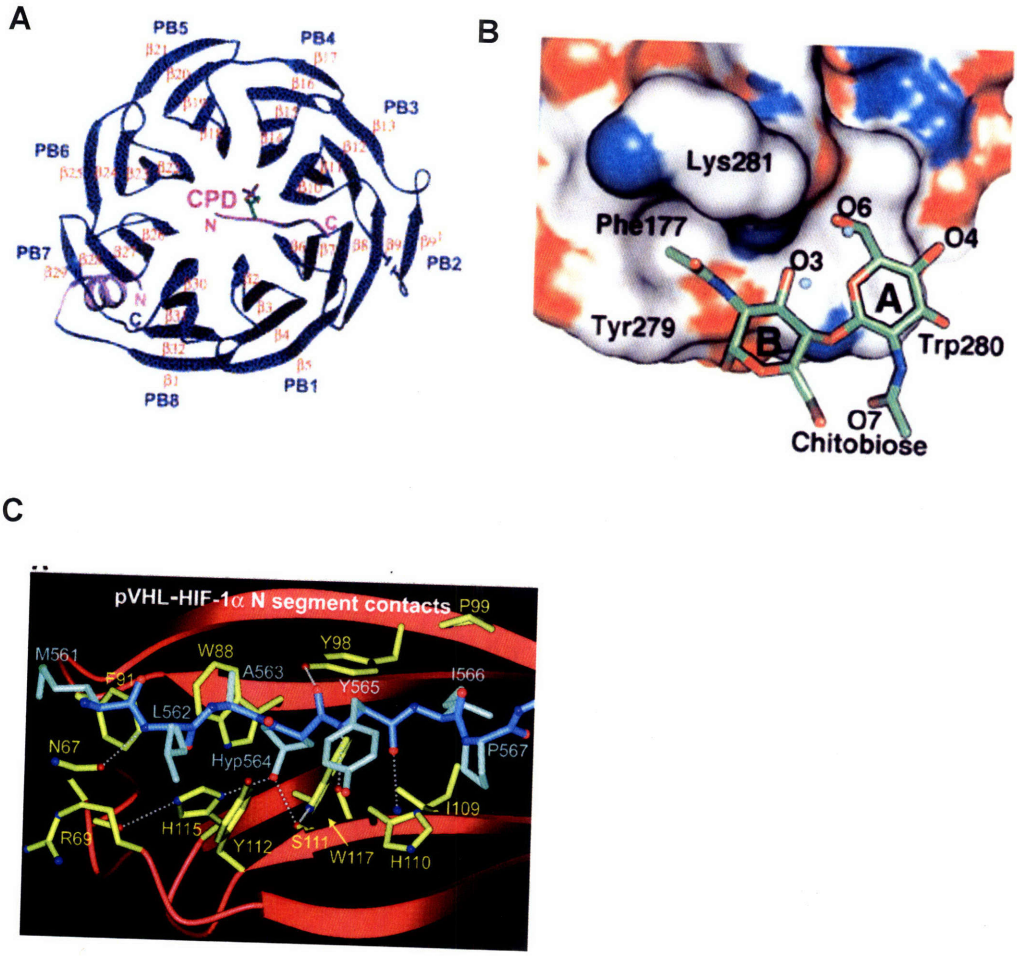


Figure 1.2. Molecular bases for recognition of eukaryotic degradation signals by cognate E3 ubiquitin ligases. (A) The SCF-Cdc4 ubiquitin ligase recognizes the Leu-Leu-phosphoThr-Pro motif on Sic1 (Orlicky et al. 2003). (B) Fbs1/Fbx2 binds to the chitobiose sugar moiety (Mizushima et al. 2004). (C) pVHL recognizes the hydroxylated Pro 564 residue on Hif-1α (Min et al. 2002).

Poly-ubiquitin chains containing at least four ubiquitin monomers are degradation signals for the eukaryotic proteasome (Thrower et al. 2000) (Fig. 1.3). Rpn10, Rad23, and Rpt5 in the 19S regulatory particle of the proteasome have been shown to bind poly-ubiquitin chains, and for Rpt5 this interaction depends on ATP hydrolysis by the ATPase subunits within this particle (Lam et al. 2002; Verma et al. 2004). Ubiquitin is removed prior to substrate degradation by deubiquitinating enzymes in order to recycle ubiquitin and possibly to save proteins carrying ubiquitin chains with fewer than four monomers from degradation (Lam et al. 1997; Nijman et al. 2005). Of the several deubiquitinating enzymes in the 19S particle, Rpn11 seems to be the primary component and loss of function mutations are lethal in yeast (Verma et al. 2002). In fact, loss of Rpn11 activity prevented degradation of a normally unstable ubiquitinated substrate, implying that deubiquitination is coupled to protein degradation by the proteasome.

In selected cases the proteasome can recognize substrates not fused to poly-ubiquitin chains. Although limited progress has been made toward determining the degradation signals on these substrates that are recognized directly by the proteasome, recent evidence points toward the requirement for an unstructured or unfolded region (Hoyt and Coffino 2004; Prakash et al. 2004). Recently, degradation of a model ubiquitin-independent substrate ornithine decarboxylase was shown to rely on both sequence elements and the specific length of its C-terminal tail (Takeuchi et al. 2008). Another substrate in this class called Cip1 uses an interaction with an alternative proteasomal lid called REGγ for degradation (Chen et al. 2007). Undoubtedly more

ubiquitin-independent substrates will surface, leading to a clearer classification of signals directly recognized by the proteasome for degradation.

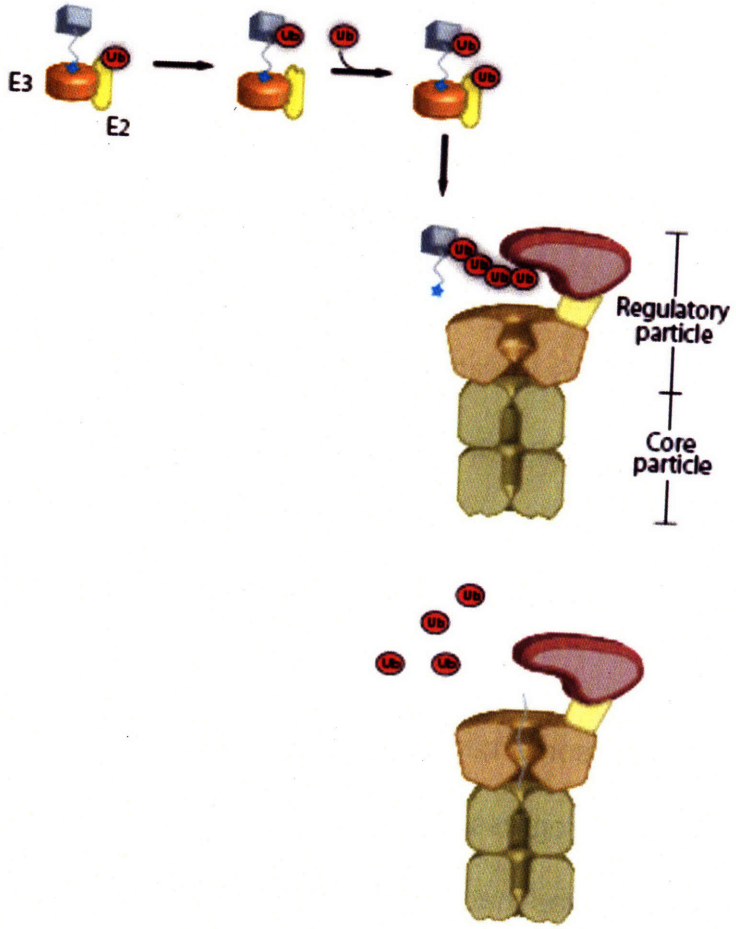


Figure 1.3. E3 ubiquitin ligases (orange) recruit proteins to E2 conjugating enzymes (yellow), resulting in poly-ubiquitinated substrates that are recognized and degraded by the 26S proteasome. The 19S lid referred to in the text is shown here as the regulatory particle, whereas the 20S proteasome is labeled as the core particle.

Prokaroytic degradation signals

Most bacterial degradation signals identified thus far are short peptide sequences. Typically these sequences are located at the termini of proteins, probably because these regions tend to be less structured and thus available to be recognized by the protease. No tagging system analogous to ubiquitin has been identified in prokaryotes or archaea, suggesting that proteolytic machines in these organisms directly recognize degradation sequences.

One of the first efforts at identifying bacterial degradation signals resulted in the isolation of hydrophobic C-terminal pentapeptides (Bowie and Sauer 1989; Parsell et al. 1990). This initial observation led to the discovery of the ssrA tag, an eleven-residue C-terminal signal that is appended to mistranslated nascent chains on stalled ribosomes (Tu et al. 1995; Keiler et al. 1996). The ssrA sequence (AANDENYALAA) is degraded by ClpAP and ClpXP (Gottesman et al. 1998), but ClpA and ClpX specifically require only the last three residues (Flynn et al. 2001) (P. Chien, unpublished observations).

Significant progress toward elucidating the diversity of ClpX degradation signals was achieved in a study isolating ClpXP substrates, yielding five classes of N- and C-terminal degradation sequences (Flynn et al. 2003). In addition to sequences that resembled the C-terminus of the ssrA tag, other complex signals were discovered. Although most substrates identified carry a putative ClpX degradation sequence, it remains to be tested whether these substrates are constitutively degraded or if access to degradation is regulated *in vivo*. For example, the zinc-binding transcription factor ZntR is degraded by ClpXP more rapidly when zinc concentrations are low or when ZntR is not bound to DNA (Pruteanu et al. 2007).

Direct interactions between the protease and substrate

ClpA must directly bind to the substrate protein to mechanically denature and translocate the polypeptide to ClpP. Recent work has elucidated regions in ClpA important for these interactions. Each ClpA hexamer was shown to bind a single ssrA peptide using a docking site in the first ATP-binding domain (Piszczek et al. 2005). A crosslinking study further refined this docking site to the loops facing the central pore of the ClpA hexamer (Hinnerwisch et al. 2005a). Substrate-binding loops are present in both ATP-binding domains of ClpA, leading to a model where the ssrA tag is transferred from the first domain to the second via these loops. Similar studies of ClpX and the ssrA tag point to a conserved theme of using pore loops to recognize degradation sequences (Siddiqui et al. 2004; Farrell et al. 2007; Martin et al. 2008).

These loops do not act in a redundant fashion to recognize the ssrA sequence, because single point mutations in any pore loop can eliminate binding of ClpA or ClpX to the ssrA tag. Currently it is also unclear how important these loops are for recognizing other degradation signals. A survey of the five classes of ClpX degradation signals raises the question of how such diverse sequences are all recognized by a single enzyme. Additionally, at any given time in the cell there may be different substrates competing for a limited number of proteases. What mechanisms can be employed by the cell to prioritize degradation of a specific class of substrates? One method is to use specificity-enhancing adaptors to alter substrate selection.

Adaptor-mediated recognition

Adaptor proteins function by binding to both the ATPase and substrate to facilitate delivery. The cellular function of adaptors is analogous to the role of E3 ubiquitin ligases because adaptors specifically recognize substrates and recruit them for degradation (Fig. 1.4). One of the first bacterial adaptors identified was RssB, encoded by the *sprE* gene, which targets the starvation response transcription factor σ^S to ClpXP for degradation during exponential growth (Muffler et al. 1996; Pratt and Silhavy 1996). The quantity of RssB in the cell is the limiting factor for σ^S degradation (Pruteanu and Hengge-Aronis 2002), and RssB activity is enhanced by phosphorylation. However, degradation of σ^S by ClpXP *in vitro* is still detectable in the absence of RssB, demonstrating that ClpX can directly recognize the degradation signal on σ^S.

The SspB adaptor was discovered by purifying a cellular factor that enhanced ssrA-tagged substrate degradation by ClpXP (Levchenko et al. 2000). Subsequent biochemical and structural evidence demonstrated that this enhancement is derived from the ability of SspB to tether ssrA-tagged proteins to ClpX, increasing the affinity and efficiency of degradation by ClpXP. As in the case of RssB, SspB is not essential for recognition and degradation of ssrA proteins by ClpXP. Interestingly, SspB does not bind to ClpA and therefore competes with ClpA for binding to the ssrA tag. SspB is therefore an example of a specificity-enhancing adaptor that shunts a class of substrates to a particular protease, therefore prioritizing degradation of that class by a particular protease.

The first adaptor for ClpA was discovered in a comparative genomic search that identified a conserved open reading frame located just 30 nucleotides upstream from

ClpA (Dougan et al. 2002). This protein, designated ClpS, binds to the N-terminal domain of ClpA and inhibits ssrA protein degradation by ClpAP. Interestingly, ClpS also prevents ClpAP from degrading model unfolded substrates like α-casein from degradation. Recently ClpS was shown to have a stimulatory function in the degradation of N-end rule substrates by ClpAP. Details of the N-end rule degradation pathway will be described in the next section, but with regard to the adaptor function of ClpS it is worthwhile to mention that ClpS greatly improves the affinity between ClpA and these substrates. ClpS thus shares the functional aspect with RssB and SspB of enhancing substrate degradation, but is unique in that it prevents ClpAP from degrading several classes of substrates.

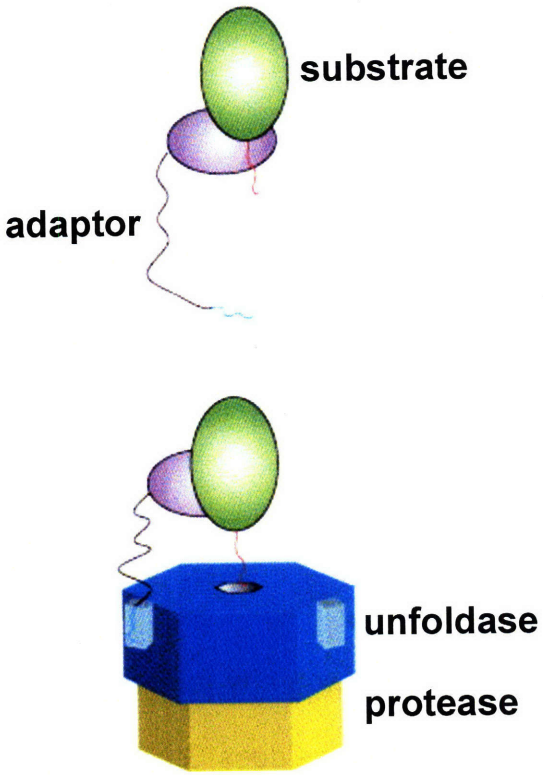


Figure 1.4. Model of adaptor-mediated substrate recognition. The adaptor uses distinct interfaces to bind substrate and dock onto the unfoldase, facilitating interactions between the unfoldase and substrate (Baker and Sauer 2006).

The N-end rule

The N-end rule correlates the *in vivo* stability of a protein to the identity of its N-terminal residue and was discovered inadvertently in a study looking at the cleavage specificity of the ubiquitin protease Ubp1 (Bachmair et al. 1986). In this study, the authors expressed ubiquitin-β-galactosidase fusion proteins in *S. cerevisiae* that were identical except for the amino acid immediately following the ubiquitin domain. After ubiquitin cleavage, the half-life of these β-galactosidase proteins varied from over twenty hours to three minutes.

Evolutionary conservation and comparison

A hierarchy exists of N-end residues in *S. cerevisiae* whereby tertiary residues like Gln and Asn are respectively converted to the secondary residues Glu and Asp by an N-terminal amidohydrolase (Nta1). N-terminal Glu and Asp residues are then recognized by the argininyl transferase Ate1, which conjugates an Arg residue to these N-termini (Varshavsky 1996). Primary N-end residues in S. cerevisiae include basic and hydrophobic residues (Fig. 1.3A), which are directly recognized by the E3 ubiquitin ligase Ubr1 and covalently linked to ubiquitin via an internal Lys residue. These poly-ubiquitinated N-end substrates are subsequently degraded by the 26S proteasome.

In addition to the primary N-end residues shown in Fig. 1.3A, mammalian systems recognize N-terminal Ala, Ser, and Thr as degradation signals. Mammalian cells also possess a mechanism to conjugate Arg to an N-terminal Cys residue that has been oxidized by nitric oxide (Kwon et al. 2002), a process that may connect N-end rule degradation to the physiological state of the cell.

In *E. coli*, only a subset of primary N-end residues is utilized relative to the repertoire of the eukaryotic N-end rule (Fig 1.3B). Basic residues are not primary destabilizing N-end residues in *E. coli* and must first be conjugated to a Phe or Leu residue at the N-terminus. This reaction is accomplished by the aminoacyl transferase Aat. The ClpAP protease is responsible for degrading N-end substrates in *E. coli* and plays a role functionally similar to that of the 26S proteasome in the N-end rule.

Preservation of the N-end rule from bacteria through mammals suggests that this pathway of protein degradation serves important physiological functions. Indeed, several processes regulated by N-end rule degradation have been discovered in eukaryotic systems and are described in the next section. It is intriguing that both *E. coli* and eukaryotes utilize aminoacyl transferases to convert secondary N-end residues to primary residues. However, this strategy likely arose independently in these two systems, because Aat and Ate1 use different substrate and recognize different N-terminal residues (Varshavsky 1996).

Physiological substrates of the N-end rule

Degradation via the N-end rule plays a role in several unrelated processes in eukaryotes. One of the first N-end substrates discovered was a G-protein subunit involved in counteracting the yeast mating pheromone response (Madura and Varshavsky 1994). Peptide import by *S. cerevisiae* is repressed by Cup1, a substrate of the N-end rule (Byrd et al. 1998). Stabilization of Cup1 in Δ*ubr1* cells decreased viability in minimal media using peptides as the sole amino acid source.

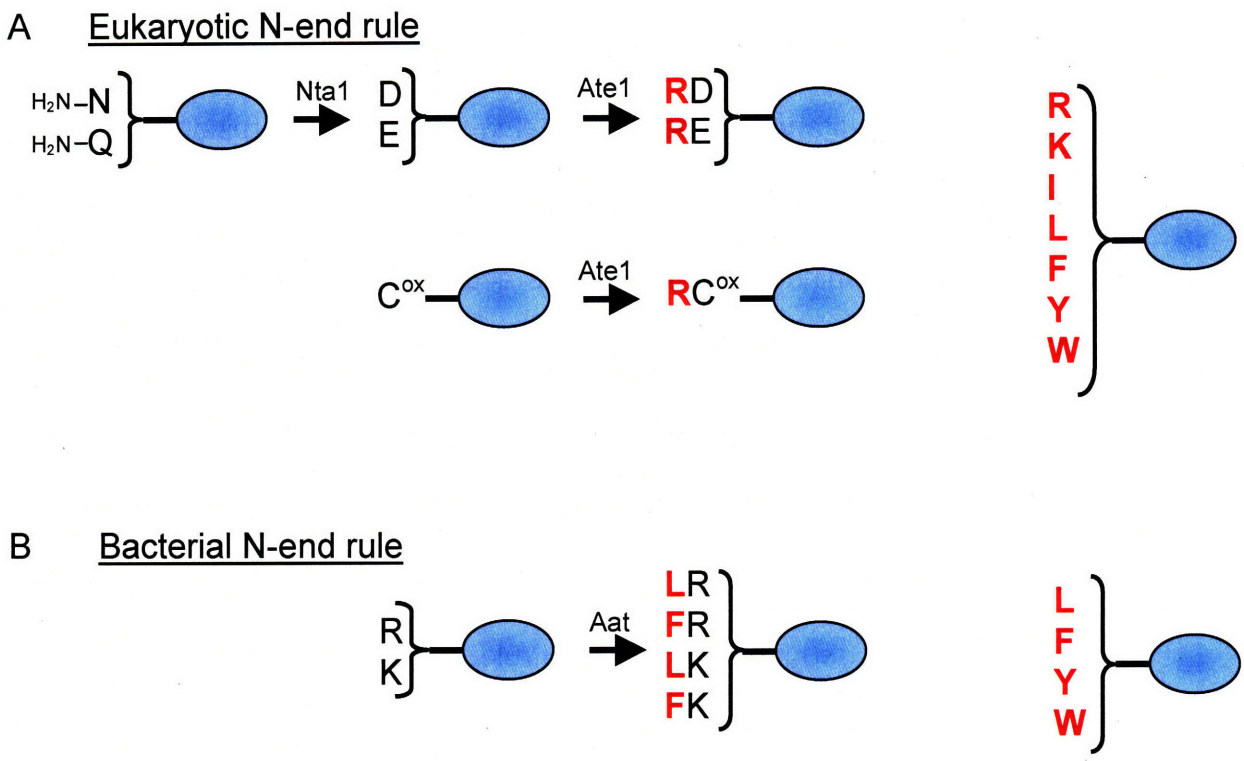


Figure 1.4. Evolutionary conservation of the N-end degradation pathway. (A) Primary N-end residues (in red) are directly recognized by the E3 ubiquitin ligase Ubr1. The eukaryotic pathway uses the amidohydrolase Nta1 to convert N-terminal N and Q residues to D and E, respectively. N-terminal D, E, and C^{ox} (oxidized Cys residue) are recognized by the argininyl transferase Atel, which conjugates an R residue to these N-termini. Residues in red are recognized by Ubr1 and targeted for ubiquitin conjugation and subsequent degradation by the proteasome. (B) N-terminal R and K residues are recognized by the aminoacyl transferase Aat, which adds an L or F residue to these N-termini. Primary N-end residues in red are recognized directly by ClpS and degraded by ClpAP.

Mammalian cardiovascular development is one striking physiological pathway that relies on the N-end rule (Kwon et al. 2002). Mice lacking the argininyl transferase Ate1 die by 17 days after fertilization with heart and blood vessel defects that ultimately result in hemorrhaging and edema. A following study identified the N-end substrates Rgs4 and Rgs5 that functioned as negative regulators of G-protein signaling involved in cardiac growth and angiogenesis (Lee et al. 2005). These regulators all carry Cys as their N-terminal residue. The prevailing model therefore hypothesizes that regulated proteolysis of these regulators promotes essential cardiovascular development during embryogenesis. The rationale for using the N-end rule pathway is thought to be due to regulated oxidation of the N-terminal Cys residue by nitric oxide, a signaling molecule in cardiac tissue (Hu et al. 2005).

Several viral proteins have also been shown to bear N-end residues in their mature protein form. Sindbis viral RNA polymerase and HIV integrase are both degraded via the N-end rule (de Groot et al. 1991; Mulder and Muesing 2000). HIV expressing integrase lacking an N-end signal replicate poorly and are defective in infection of human culture cells (Lloyd et al. 2007).

To date no endogenous bacterial N-end substrates have been identified, complicating the progress of understanding the importance of the N-end rule in bacteria and the diversity of N-end signals. The isolation of N-end substrates will elucidate cellular processes dependent on the N-end rule for regulation.

Generation of N-end rule signals

A central question of the N-end rule is the mechanism by which N-end signals are revealed. Nascent polypeptide chains are initiated with a Met residue (formyl-Met in prokaryotes) that can be removed by methionyl aminopeptidases if the second residue is small and uncharged (Flinta et al. 1986). As a result, N-end residues adjacent to the initiator Met residue are not usually exposed by aminopeptidases. The one notable exception is Cys, which is the second residue of Rgs4 and Rgs5, the two aforementioned regulators of cardiac growth. In these cases, the N-terminal Cys residues must be oxidized prior to arginylation and therefore removal of the N-terminal Met residue does not immediately target these polypeptides for N-end rule processing.

The examples of viral N-end substrates described above are produced by viral proteases that cleave multidomain polypeptide chains. In the case of Sindbis RNA polymerase, the viral protease cleaves N-terminal of the naturally occurring Phe residue but can tolerate residues in this position as different as Met, Ala, and Arg (de Groot et al. 1991). This proteolytic promiscuity is also true of HIV protease (Mulder and Muesing 2000), but since either a Trp or Phe residue is located at the analogous position in other lentiviral integrases, N-end rule degradation of the integrase may be a general regulatory mechanism for the proliferation of lentiviruses.

Similar proteolytic processing mechanisms are used for the generation of eukaryotic N-end substrates. Cohesin degradation was mentioned earlier as a critical step leading to chromosome segregation during anaphase. The cysteine protease Esp1 cleaves cohesin to reveal an N-terminal Arg (Rao et al. 2001). In turn, Esp1 is regulated by an inhibitor called securin, which is ubiquitinated and degraded at the onset of

anaphase (Peters 2002). In a similar fashion, the inhibitor of apoptosis factor DIAP is cleaved by an unidentified caspase to reveal an N-terminal Asn, a tertiary N-end residue (Ditzel et al. 2003).

Model N-end substrates can be generated both *in vivo* and *in vitro* using a protease that tolerates or requires N-end residues at the position adjacent to the cleavage site. The Varshavsky lab has used Ubp1 to remove N-terminal ubiquitin domains from fusion proteins because this protease tolerates every residue except for Pro at the position immediately following ubiquitin. Using the same logic, the SUMO protease Ulp1 efficiently cleaves the SUMO domain from a fusion protein. Ulp1 recognizes the tertiary structure of SUMO and cleaves after a di-glycine motif unless the following residue is Pro (Li and Hochstrasser 1999). Another strategy to generate N-end substrates *in vitro* is the use of enterokinase protease, which requires the pentapeptide sequence Asp_4-Lys preceding the cleavage site but tolerates most residues in the position immediately following the site (Cranz-Mileva et al. 2008).

Recognition components of the N-end signal

Ubr1 is the 225 kilodalton E3 enzyme that recognizes primary N-end residues in yeast (Bartel et al. 1990). It contains two distinct binding sites for basic and hydrophobic N-end residues (Baker and Varshavsky 1991). To promote ubiquitination of N-end substrates, Ubr1 interacts with the E2 conjugating enzyme Ubc2 (Dohmen et al. 1991). No structural information exists on the molecular basis of N-end selectivity by Ubr1 or the architecture of the binding sites, although the two binding sites for basic and hydrophobic N-end residues are both located in the first 1170 residues of Ubr1 (Du et

al. 2002). Seven paralogues of Ubr1 are thought to exist in mammals, though not all of these proteins interact with N-end signals (Tasaki et al. 2005; Tasaki et al. 2007).

The initial discovery that ClpAP was the sole protease responsible for degrading N-end substrates implied that ClpA directly recognized these degradation sequences. However, reconstitution of efficient N-end degradation in vitro required the ClpS adaptor, and deletion of the *clpS* gene appeared to abrogate turnover of an N-end β-galactosidase substrate *in vivo* (Erbse et al. 2006). ClpS protein interacted directly with peptides on a peptide array bearing an N-end residue (Phe, Tyr, Trp, Leu), but binding was reduced for peptides with acetylated α-amino groups or that contained multiple acidic residues (Erbse et al. 2006).

ClpS exhibits limited homology in its protein sequence with Ubr1, leading Erbse and coworkers to mutate two conserved Asp residues to Ala and successfully diminish degradation of an N-end green fluorescent protein (GFP) by ClpAP and ClpS. This result suggested that the N-end signal binding site on ClpS contains or is adjacent to $Asp^{35}Asp^{36}$. Indeed, Asp^{36} is one component of a hydrogen bonding network that contacts the α-amino group of the N-terminal residue, although Asp^{35} interacts only through its backbone carbonyl oxygen (chapter 4). Overall, residues in ClpS that contact the N-terminal residue are well-conserved in bacterial orthologues of ClpS and in eukaryotic Ubr1 sequences, suggesting evolutionary conservation of N-end signal recognition.

My investigation of the bacterial N-end rule was initated as a result of a selection for N-terminal degradation signals. Out of about 300 sequences tested, more than 30 were verified as N-terminal signals and all were categorized as N-end rule sequences.

Degradation of proteins bearing these N-end signals *in vivo* depended on ClpA but not ClpS, although the presence of ClpS accelerated N-end substrate degradation. This effect of ClpS is likely due to the ability of ClpS to enhance the affinity of ClpAP for N-end substrate from 7 to 70-fold *in vitro*.

ClpAP in the presence of ClpS specifically degraded a model substrate bearing any of the four N-end residues *in vitro*, but the efficiency of degradation was reduced by acidic residues near the N-terminus. ClpS bound poorly to these acidic N-end sequences, but ClpAP in the absence of ClpS was not similarly affected. However, ClpAP was unable to degrade substrates whose N-end signals were located too close to the folded region of the substrate even though ClpS bound these substrates efficiently. These results indicate that ClpS and ClpAP are both important for efficient N-end degradation but possess distinct sequence requirements.

ClpS possessed the hallmarks of N-end rule selectivity in that it bound to peptides with N-terminal residues Phe, Tyr, Trp, and Leu but not other hydrophobic residues (Erbse et al. 2007; Wang et al. 2008). To gain insight into the mechanism by which ClpS specifically recognizes an N-end residue only at the N-terminal position, the cocrystal structure of *C. crescentus* ClpS in complex with a peptide containing an N-terminal Tyr residue was solved. ClpS uses a hydrogen bonding network to recognize the α-amino group, a unique feature of the first residue of a protein. The N-terminal Tyr side chain is buried in a pocket that is lined with hydrophobic side chains in ClpS, and mutation of a single residue in this pocket alters the substrate selectivity of ClpS. Taken together, my work has elucidated sequence and structural requirements for efficient recognition of the N-end signal by ClpAP and ClpS.

Chapter Two

ClpS modulates but is not essential for bacterial N-end rule degradation

This chapter was published as Wang, K.H., Sauer, R.T., and Baker, T.A. 2007. *Genes & Dev.* **21:** 403-408. R.T. Sauer and T.A. Baker assisted in experimental design and preparation of the manuscript.

Abstract

In eubacteria, the ClpS adaptor has been proposed to be essential for degradation of N-end rule substrates by the AAA+ protease ClpAP. To test this model, we assayed degradation of substrates bearing N-end rule sequences isolated in a genetic screen for efficient degradation tags. ClpS was not vital for degradation *in vivo* but rather stimulated turnover in a sequence-specific manner. Although ClpS substantially enhanced degradation of N-end substrates at low substrate concentrations *in vitro*, it suppressed the degradation rate when substrate was saturating. Thus, we conclude that ClpAP recognizes N-end rule substrates directly, whereas ClpS modulates this degradation pathway.

Introduction

Protein degradation regulates many biological processes as well as protein quality control in both prokaryotes and eukaryotes. The ability of proteolytic enzymes to recognize their substrates with precise specificity is important in carrying out efficient degradation of desired substrates and avoiding unwanted destruction of essential cellular proteins. The 26S proteasome is the primary degradation complex in eukaryotes. In *E. coli*, there are five ATP-dependent proteases: ClpAP, ClpXP, HslUV, Lon, and FtsH (Schirmer et al. 1996). Some proteolytic substrates such as ssrA-tagged polypeptides are degraded by multiple proteases, whereas other proteins are degraded exclusively by one protease (Herman et al. 1995; Gottesman 1996; Gottesman et al. 1998).

The AAA+ ATPase subunits in the 26S proteasome and bacterial proteases recognize specific degradation signals on substrates, unfold these substrates, and then translocate the denatured polypeptides into a proteolytic chamber where degradation occurs (Singh et al. 2000; Benaroudj et al. 2001). Degradation signals vary in structure and complexity and include phosphorylated residues, polyubiquitin chains, and exposed peptide sequences (Parsell et al. 1990; Pickart 1997; Craig and Tyers 1999; Pickart 2001). In bacteria, the best-characterized degradation tags are simple peptide sequences that interact directly with the protease (Gottesman et al. 1998; Kim et al. 2000). Proteomic studies have identified several classes of sequence motifs that target proteins for ClpXP degradation (Flynn et al. 2003), but the mechanisms that mediate substrate recognition by any of the intracellular bacterial proteases are still not thoroughly understood.

A particularly interesting mode of proteolytic recognition is the N-end rule, which states that the stability of a protein *in vivo* depends on the identity of its N-terminal residue

(Bachmair et al. 1986). This degradation pathway is present in both eukaryotes and prokaryotes (Varshavsky 1992). Some important eukaryotic N-end rule substrates include Scc1, a protein involved in maintaining chromosomal cohesion (Rao et al. 2001), and the γ-subunit of a bovine neural G-protein (Hamilton et al. 2003). Natural N-end substrates have not been characterized in prokaryotes, but experiments with model substrates reveal that proteins with Leu, Tyr, Trp, or Phe at the N-terminus are degraded in *E. coli* if the ClpAP protease is present (Tobias et al. 1991). Moreover, proteins that initially have Arg or Lys at the N-terminus are converted into N-end rule substrates by addition of an N-terminal Phe or Leu in a reaction catalyzed by the Aat aminoacyl transferase (Tobias et al. 1991).

The ClpAP protease is composed of ClpA, a hexameric AAA+ ATPase and protein unfoldase, and ClpP, a 14-subunit compartmental protease (Katayama et al. 1988; Maurizi et al. 1990; Weber-Ban et al. 1999). ClpS is a monomeric adaptor protein that binds to the N-terminal domain of subunits in the ClpA hexamer and alters substrate recognition by ClpAP (Dougan et al. 2002; Xia et al. 2004). For example, ClpS inhibits ClpAP degradation of ssrA-tagged substrates but enhances degradation of other substrates. A recent study has reported that ClpS binds directly to the N-terminal destabilizing residues of N-end rule substrates and is essential for their degradation by ClpAP (Erbse et al. 2006). This model predicts that N-end rule substrates should not be recognized or degraded by ClpAP in the absence of ClpS.

To explore the bacterial N-end rule system in greater detail, we devised a genetic screen for N-terminal sequences that direct degradation of a conditionally toxic *E. coli* protein and isolated a collection of substrates with varied degradation signals. Examination of the cellular levels of these N-end rule substrates revealed that ClpS is not required for N-end degradation *in vivo*, although proteolysis occurred more rapidly

when this adaptor was present. To determine if ClpAP recognizes N-end rule substrates directly, we constructed and purified proteins bearing different N-end rule tags and measured degradation rates *in vitro*. In the absence of ClpS, ClpAP degraded each of these substrates but not a parental control protein. This result indicates that ClpA has a receptor for N-end rule sequence motifs. ClpS enhanced ClpAP degradation of low concentrations of N-end rule substrates but reduced degradation of high concentrations of these substrates. Thus, ClpS improves binding of N-end rule substrates to ClpAP but slows the maximal degradation rate. We also found that sequences beyond the N-terminal residue affected recognition, demonstrating that N-end rule degradation is influenced by more than one residue. In contrast to previous studies, our results indicate that ClpA has an intrinsic ability to recognize and degrade N-end rule substrates. Rather than being essential for this process, ClpS modulates substrate recognition and turnover.

Results and Discussion

Isolation of strong N-terminal degradation signals

To identify sequences that could act as N-terminal degradation signals, we constructed a screening vector that expressed a fusion protein consisting of the SUMO domain from *S. cerevisiae*, five randomized residues, the conditionally toxic mPheS enzyme, and a C-terminal Flag epitope (Fig. 2.1A). The mPheS enzyme is a mutant of *E. coli* phenylalanyl tRNA synthetase that allows incorporation of chloro-phenylalanine (Cl-Phe) into proteins and thus kills cells grown in the presence of this amino-acid derivative (Kast and Hennecke 1991). The screening vector was transformed into cells expressing the Ulp1 protease (Li and Hochstrasser 1999) to remove the SUMO domain and reveal the randomized sequence at the N-terminus of mPheS.

We found that cells expressing mPheS with roughly 35 different N-terminal sequences displayed comparable growth on media with or without 15 mM Cl-Phe and slow growth on 20 mM Cl-Phe. By contrast, cells expressing most mPheS variants did not grow in the presence of either concentration of Cl-Phe. Moreover, clones that grew well in the presence of 20 mM Cl-Phe had frameshift or termination mutations in the randomized segment of the fusion protein (data not shown). For mPheS proteins with N-terminal sequences that allowed growth in the presence of Cl-Phe, we assayed intracellular stability by western blotting at different times after blocking protein synthesis with spectinomycin. Six of the original 35 clones did not display detectable degradation of mPheS and were discarded. The 29 sequences listed in Fig. 2.1B caused at least 75% turnover over the course of 1 h, although degradation mediated by some sequences was substantially faster. For example, mPheS proteins with YIALR or YLFVQ at the N-terminus were completely degraded after 30 min, whereas a protein with LVKEL at the N-terminus showed only modest turnover at this time point (Fig. 2.1C). In contrast, two

substrates that did not confer survival in the presence of Cl-Phe were not significantly degraded over 1h (data not shown).

All of the N-terminal sequences identified by the criteria discussed above began with Phe, Leu, Arg, Trp, or Tyr (Fig. 2.1B). Each of these amino acids has previously been identified as an N-end rule residue in *E. coli* (Tobias et al. 1991), suggesting that our screen resulted in the isolation of N-end rule substrates. We did not recover sequences beginning with Lys, which is also a bacterial N-end residue, but our total library was small (≈ 350 clones) and contained only a tiny fraction of the possible five-residue N-terminal sequences. To determine if these sequences function as degradation signals only when present at the N-terminus, we tested the intracellular stability of the YLFVQ-mPheS variant in cells with or without Ulp1 (Fig. 2.1C). In the absence of Ulp1, the SUMO-YLFVQ-mPheS fusion was stable over 1h, indicating that the YLFVQ sequence needs to be exposed at the N-terminus to function as an efficient degradation signal.

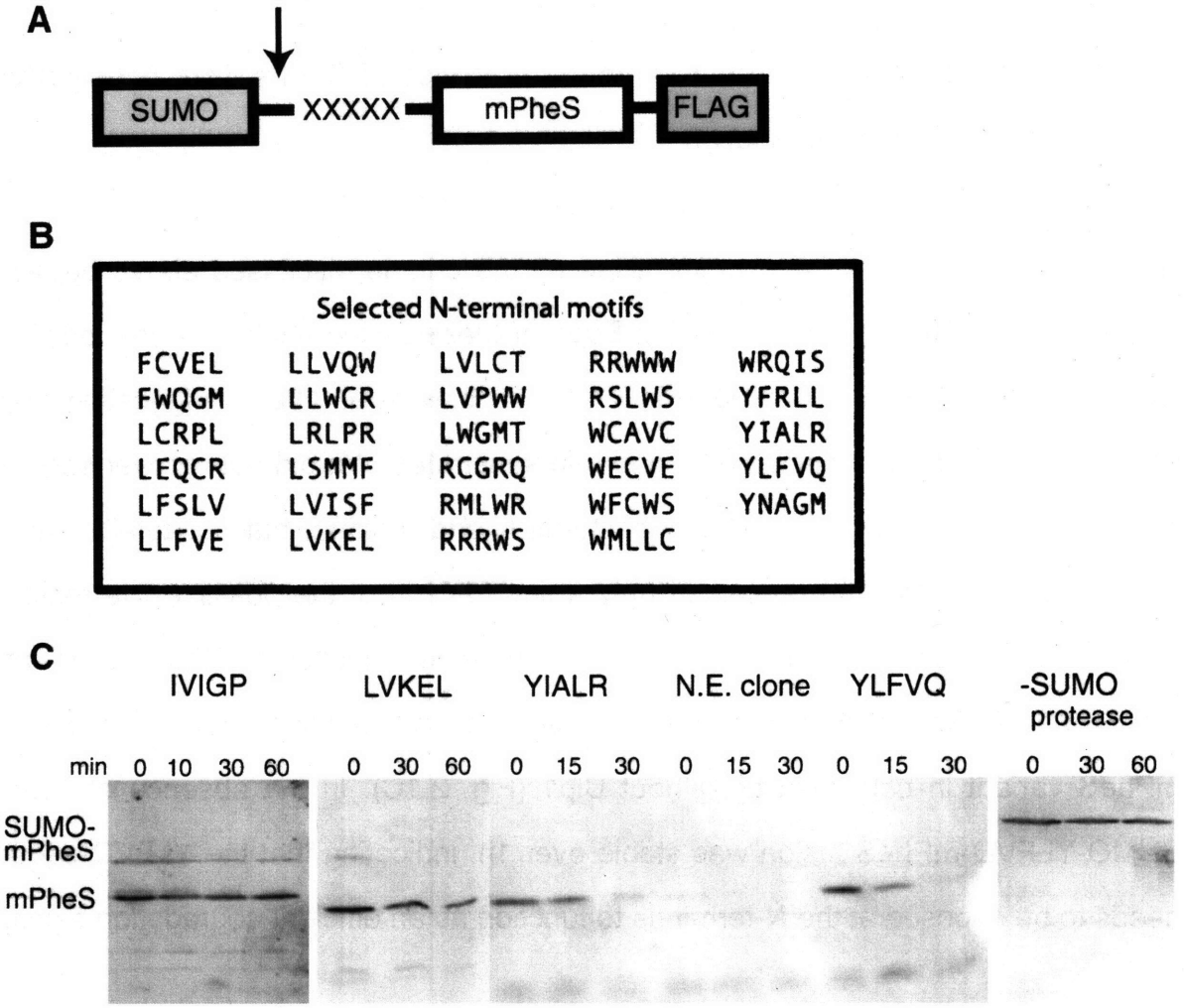


Figure 2.1. The mPheS selection assay yields N-end degradation signals of varying strengths. (*A*) The SUMO-mPheS protein used for the selection contains five randomized residues which become the N-terminal amino acids of mPheS following cleavage by the Ulp1 SUMO protease (indicated by the arrow). (*B*) List of N-terminal tags that passed the selection and directed degradation *in vivo*. (*C*) Turnover of selected mPheS variants in wild-type cells was assayed at different times (in minutes) after stopping translation with spectinomycin. mPheS protein was detected by western blotting using α-FLAG antibody. The sample lacking SUMO protease expressed SUMO-YLFVQ-mPheS. N.E., a non-expressing clone due to a frameshift in the N-terminal sequence.

ClpA but not ClpS is a required N-end rule component in vivo

To test the importance of ClpA and ClpS in intracellular degradation, mPheS proteins with WFCWS or WECVE sequences at the N-terminus were introduced into strains lacking one or the other of these N-end rule components. Consistent with the observation that ClpA is an essential component of bacterial N-end rule degradation (Tobias et al. 1991), Δ*clpA clpS*$^+$ cells expressing either substrate failed to grow on plates containing 15 mM Cl-Phe (Fig. 2.2A). Surprisingly, however, both the *clpA*$^+$ *clpS*$^+$ and *clpA*$^+$ Δ*clpS* cells containing these substrates grew well in the presence of Cl-Phe (Fig. 2.2A,B). Plating serial dilutions of the two strains revealed that the Δ*clpS* cells had a mild plating defect on Cl-Phe, which was expressed principally as smaller colonies but also resulted in an approximately twofold to threefold reduced plating efficiency with some mPheS variants. However, the phenotype of the Δ*clpS* cells was clearly much milder than that of the Δ*clpA* strain. Thus, these data suggest that ClpS is less important than ClpAP in clearing N-end rule substrates from the cell.

ClpS stimulates N-end degradation in vivo in a sequence-specific manner

To compare the degradation rates of mPheS proteins with N-terminal WFCWS and WECVE sequences, we measured the half-lives of these substrates in liquid culture. The WFCWS-mPheS protein had a half-life of 5 min in *clpA*$^+$ *clpS*$^+$ cells, 43 min in *clpA*$^+$ Δ*clpS* cells, and 83 min in Δ*clpA* *clpS*$^+$ cells (Fig. 2.2C). The WECVE-mPheS protein was degraded with half-lives of 24 min, 40 min, and 122 min in these strains (Fig. 2.2D). Because both proteins begin with the same N-end rule residue (Trp) and yet were degraded at rates differing by almost fivefold in *clpA*$^+$ *clpS*$^+$ cells, residues past the N terminus appear to contribute to recognition of these substrates. Moreover, ClpS enhanced ClpAP degradation of both proteins, albeit to different extents depending on the degradation signal (approximately eightfold for WFCWS and twofold for WECVE). Hence, the stimulatory effect of ClpS also depends on more than the N-terminal residue. Finally, elimination of ClpA slowed degradation more than elimination of ClpS in both cases. This result is consistent with experiments from the previous section, which indicated that ClpA is more important than ClpS for the phenotype associated with intracellular degradation of N-end rule substrates.

The results presented so far support the idea that ClpAP is the protease responsible for N-end rule degradation in *E. coli* (Tobias et al. 1991). Importantly, however, ClpS appears to be an auxiliary factor that enhances but is not essential for this degradation. In the following sections, we present the results of in vitro experiments that support these conclusions.

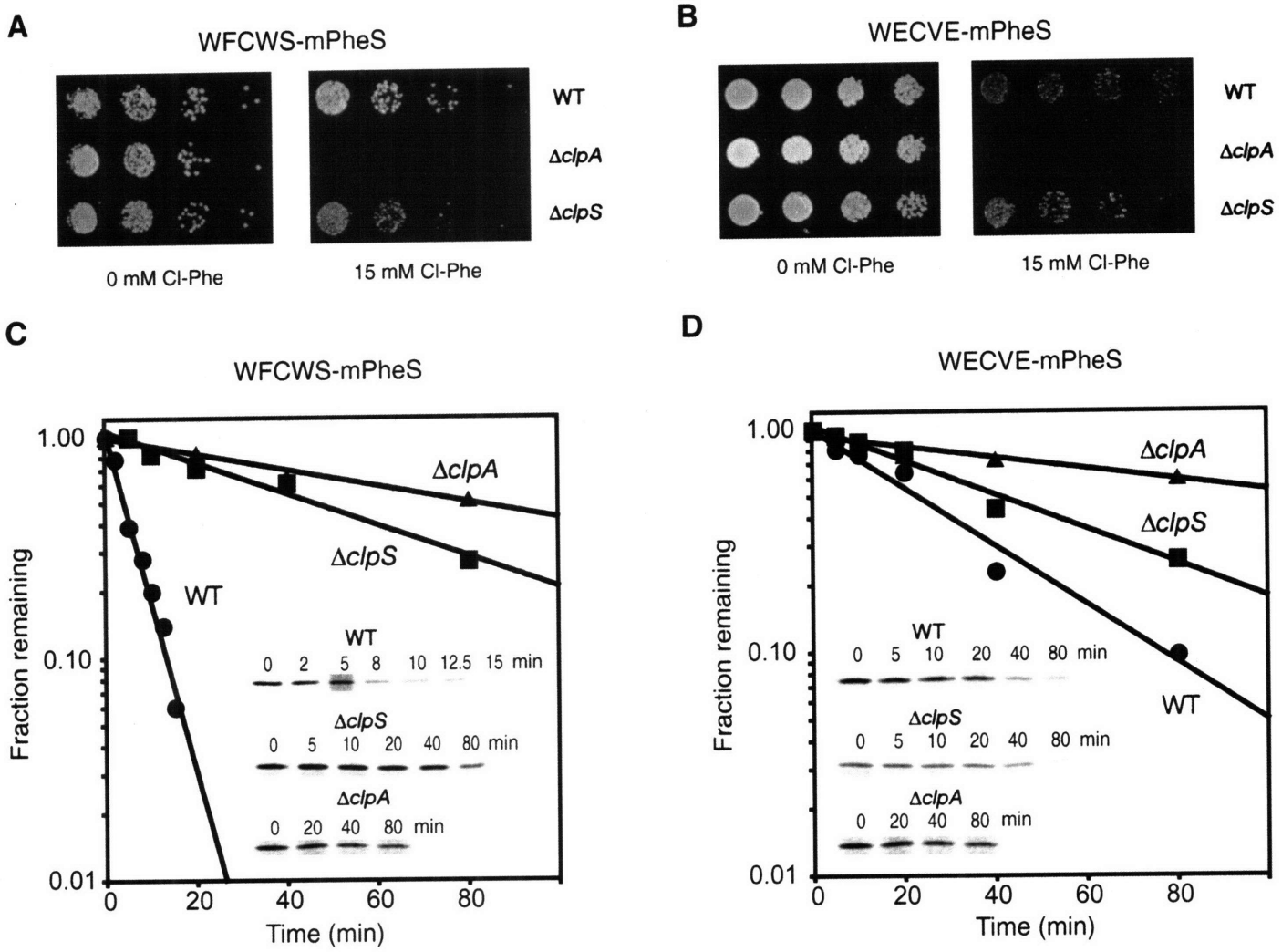


Figure 2.2. Effects of ClpS and ClpA on activity, steady-state levels, and degradation of N-end rule substrates *in vivo*. Wild-type, Δ*clpA*, or Δ*clpS* cells expressing SUMO protease and (*A*) WFCWS- or (*B*) WECVE-mPheS were serially diluted and replica plated on 0 and 15 mM Cl-Phe. (*C*) Quantification of pulse-chase experiments for ^{35}S-labelled WFCWS-mPheS in wild-type, Δ*clpS*, and Δ*clpA* strains. The half-lives were 5,

43, and 83 minutes, respectively. (*D*) Quantification of WECVE-mPheS degradation in the same three strains gave half-lives of 24, 40, and 122 min.

ClpAP has a receptor site for N-end motifs

For degradation experiments *in vitro*, we choose five of the N-terminal tags shown in Fig. 2.1B and cloned these sequences between a His_6-tagged SUMO domain and the I27 domain of the muscle protein titin (His_6-SUMO-X_5-titin). This fusion protein was purified by Ni^{++}-NTA chromatography, cleaved with Ulp1, repurified by passage through Ni^{++}-NTA to remove uncleaved fusion protein and the His_6-SUMO fragment, and subjected to a final gel-filtration purification step.

We found that purified ClpAP degraded 4 μM YFRLL-titin in the absence of ClpS and that proteolysis was accelerated threefold in the presence of ClpS (Fig. 2.3A). As a control for the purity of ClpA, we assayed for but did not detect any ClpS contamination by western blotting (data not shown).

Moreover, degradation was N-end rule dependent because ClpAP did not degrade TMCMK-titin, which is not an N-end substrate, over the course of 1 h (Fig. 2.3A inset). When degradation of 40 μM YFRLL-titin was assayed, similar rates of ClpAP proteolysis were observed whether ClpS was present or absent (Fig. 2.3B). These results demonstrate that ClpAP degrades an N-end rule substrate in the absence of ClpS. Additionally, because the stimulatory effect of ClpS was observed at low but not high substrate concentrations, it appears that this adaptor acts to increase the affinity of ClpAP for N-end rule substrates.

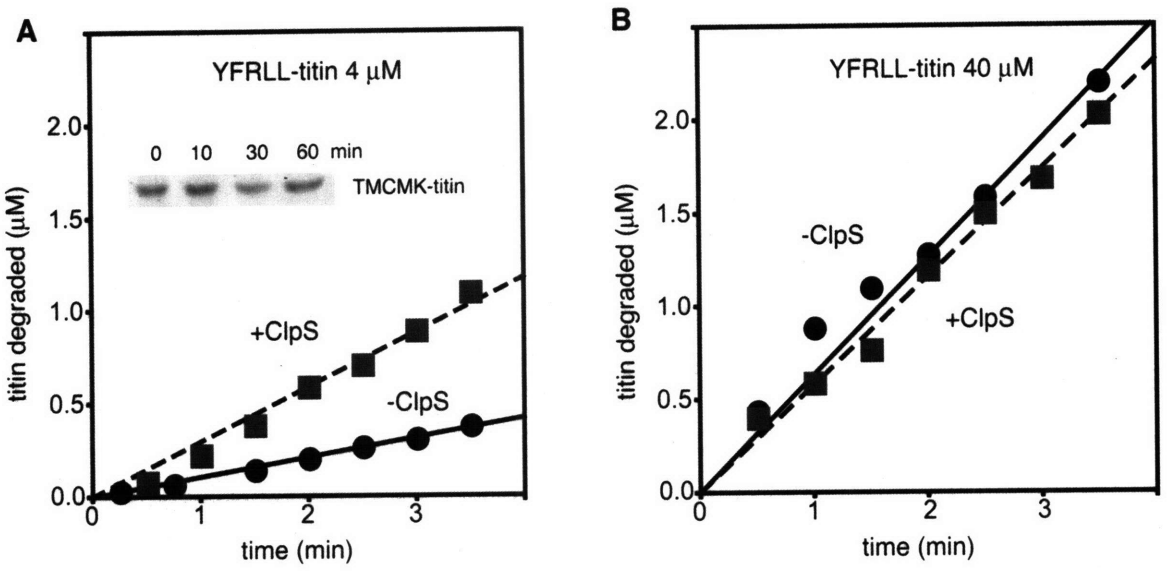


Figure 2.3. ClpS improves but is not required by ClpAP for N-end degradation *in vitro*. Degradation of YFRLL-titin was followed by measuring the release of ^{35}S-labelled acid-soluble peptides at different times after addition of ATP. The concentration of substrate was (*A*) 4 μM or (*B*) 40 μM. (*inset*) TMCMK-titin (10 μM) was not degraded by ClpAP plus ClpS. The TMCMK sequence fused to mPheS did not pass the Cl-Phe selection Protein was detected with Sypro orange stain (Molecular Probes).

ClpS enhances substrate affinity but compromises enzymatic turnover

To determine whether ClpS enhances ClpAP-mediated degradation of N-end substrates by strengthening enzyme•substrate interactions, we used Michaelis-Menten analysis to determine K_M and V_{max} values for titin substrates with N-terminal WECVE, WFCWS, YFRLL, YLFVQ, and LLWCR signals (Fig. 2.4; Table 2.1). In each case, degradation was observed in the absence of ClpS, but the presence of the adaptor decreased K_M by factors ranging from 10-fold to 70-fold. This reduction in K_M allowed faster degradation at low substrate concentrations. Hence, the principal effect of the ClpS adaptor is to strengthen binding of N-end rule substrates to ClpAP when substrate is scarce. Because the identity of the degradation tag influenced the magnitude of the ClpS-dependent reduction in K_M, sequence context must play some role in this process.

Unexpectedly, ClpAP degraded high concentrations of each of the five N-end rule titin substrates significantly faster when ClpS was absent (Fig. 2.4). Indeed, values of V_{max} were 1.8-fold to 3.5-fold higher in the absence of ClpS compared to its presence (Table 1). This inhibitory effect of ClpS on V_{max} is not a general property of adaptor proteins. For example, the SspB adaptor decreases K_M and increases V_{max} for ClpXP degradation of cognate substrates (Levchenko et al. 2000; Wah et al. 2003; Flynn et al. 2004; McGinness et al. 2006). It is possible that ClpS binding to ClpA simply reduces the maximal rate at which the enzyme can unfold or translocate substrates. Alternatively, slower maximal degradation might reflect inefficient handoff of the N-end rule substrate from ClpS to ClpA. Erbse et al. (2006) have shown that ClpS interacts with the α-amino group and the side chain of the N-terminal residue of N-end rule substrates. Because our results indicate that ClpA is likely to recognize these same determinants, it is possible that dissociation of the N-end rule substrate from ClpS becomes the rate-limiting step for degradation by ClpAP at saturating substrate concentrations.

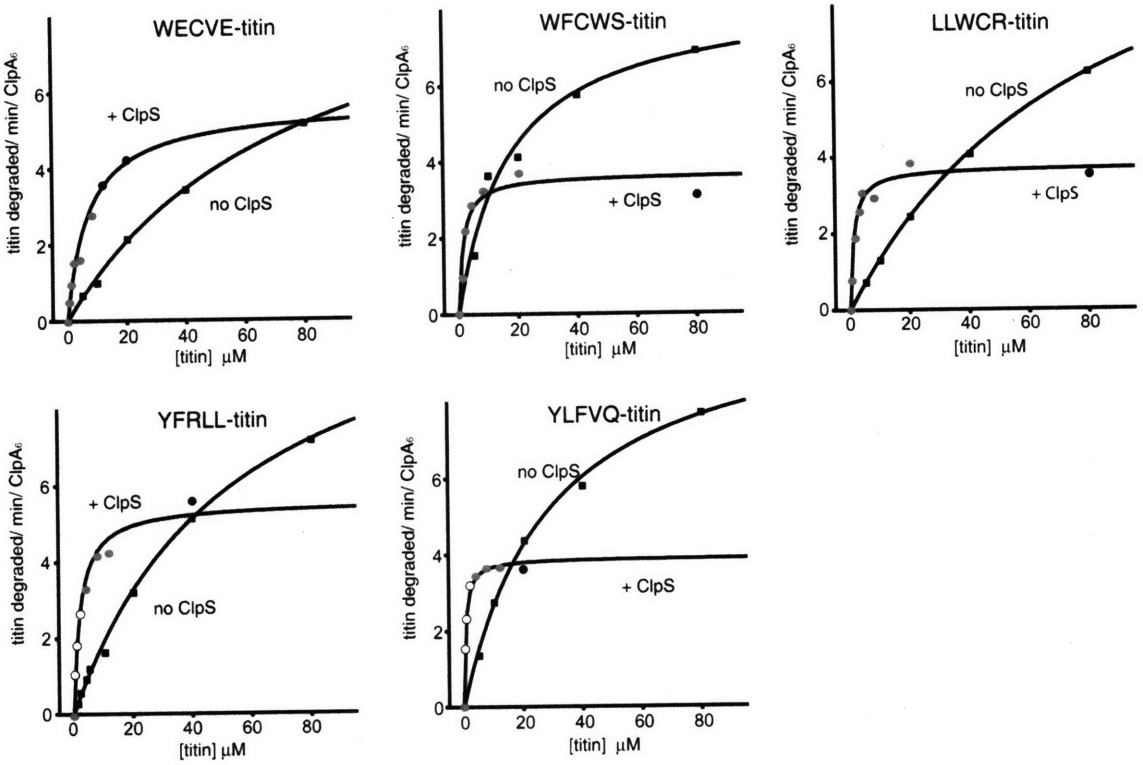


Table 1. *Degradation rates for titin substrates by ClpAP in vitro*

	No ClpS		+ClpS	
N-end motif	K_M	V_{max}	K_M	V_{max}
LLWCR	87	13	1.2	3.7
WECVE	78	10	7.3	5.7
WFCWS	17	8.4	1.5	3.7
YFRLL	58	12	2.4	5.5
YLFVQ	29	11	0.6	3.9

K_M constants are in units of micromolar; V_{max} units are titin degraded/min/$ClpA_6$.

Figure 2.4. ClpS enhances the affinity of ClpA for N-end substrates but suppresses the maximal rate of degradation. The initial rates of degradation per $ClpA_6$ in the presence (●) or absence (■) of ClpS were measured as a function of substrate concentration. $ClpA_6$ concentrations were used at 25(○), 50(◉), or 100(●) nM (see Methods), and ClpS was present at a ratio of 9 monomers per $ClpA_6$. Each data point was the average of at least two separate kinetic experiments, and the error for each point was between 5 and 10 % of the average value.

Strength of N-end rule degradation signals

In its simplest form, the N-end rule states that the half-life of a protein *in vivo* is determined by the identity of its N-terminal residue (Bachmair et al. 1986). Our results support the idea that specific N-terminal residues are required for N-end rule degradation in bacteria but also suggest a greater degree of signal complexity, as we find that ClpAP degraded substrates with identical N-terminal residues at substantially different rates *in vivo* and *in vitro* (Fig. 2.2; Fig. 2.4). A simple amendment to the N-end rule would be that the identities of the residues immediately following the N-terminal residue also influence the strength of interaction of the degradation tag with ClpAP and/or ClpS. In other words, an N-end rule degradation tag includes residues in addition to the N-terminal amino acid. It remains to be determined how residues at positions other than the N-terminus influence the degradation of N-end rule substrates by ClpAP, although Erbse *et al.* (2006) have suggested that a net positive charge near the N-terminus may be important for ClpS binding. We note, however, that many of our N-end rule degradation tags contain no basic residues (Fig. 2.1B), and that tags with a strong net negative charge (like WECVE) can function as efficient degradation signals. Systematic mutagenesis of the signals on some of the substrates presented here is in progress to address the role of sequence context of recognition by ClpAP and ClpS.

The bacterial N-end rule degradation pathway

The results presented here contradict the recent proposal that ClpS is an essential component of the N-end rule degradation pathway in bacteria (Erbse et al. 2006). We find that ClpAP can recognize and degrade N-end rule substrates in the absence of ClpS, both *in vitro* and *in vivo*. Moreover, although ClpS enhances ClpAP degradation of low concentrations of N-end rule substrates, it suppresses degradation when substrate is abundant. Thus, depending upon conditions, the ClpS adaptor can function either as an enhancer or inhibitor of ClpAP-mediated degradation of N-end rule substrates.

Natural N-end rule bacterial substrates will need to be identified before it will be possible to determine whether ClpS plays a major biological role, either positively or negatively, in the degradation substrates with this class of degradation signal.

Our studies show that ClpS is a modifier rather than an essential component of N-end rule degradation and raise numerous questions about mechanism. How does ClpS enhance substrate binding of N-end rule substrates to ClpA if both the adaptor and protease interact with the same or overlapping determinants in the substrate? Does reversible binding of N-end rule substrate to ClpS, which in turn binds to each subunit of the ClpA hexamer, simply keep the local concentration of these substrates high, or is the substrate handed off directly from the adaptor to the enzyme? Is the way in which ClpS inhibits ClpAP degradation of ssrA-tagged substrates related to the mechanism by which it slows proteolysis of high concentrations of N-end rule substrates? Biochemical experiments employing some of the N-end rule signals identified here should help answer many of these questions.

Materials and Methods

Strains and plasmids

E. coli K12 strain W3110 was used for all experiments *in vivo*. W3110 *clpA::kan* and W3110 Δ*clpS* were kindly provided by J. Hou (M.I.T). All proteins were purified from the BL21 strain of *E. coli* carrying the λDE3 T7 polymerase lysogen and the appropriate overexpression vector (see below).

The library plasmid was constructed from the vector pKSS (Kast 1994) by inserting the *smt3* gene encoding SUMO using *Hind*III/*Bam*HI sites. Codons 96-97 were altered from ATTGGT to ACCGGT to insert an *Age*I site, causing an amino-acid change that did not prevent efficient cleavage by SUMO protease. A new Shine-Dalgarno sequence was inserted to replace the sequence lost during insertion of *smt3* (TAAGGAGGTtaacaATG), and site-directed mutagenesis was performed to remove two Shine-Dalgarno sequences within *smt3* (aaa aga→AAA CGT and atg gag gat→ATG GAT GAT).

mPheS in the pKSS vector was replaced with a version containing a *Sac*I site at codons 6-7 (GAACTG to GAGCTC) and a C-terminal FLAG epitope tag. The randomized cassette was constructed from an oligonucleotide containing five repeats of N-N-G/C (where N represents A, T, G, or C) flanked by *Age*I and *Sac*I sites. This randomization scheme allows all 20 amino acids and one stop codon but should reduce the bias between amino acids encoded by multiple codons and those encoded by one codon. Nevertheless, biases in our population were evident, including an abundance of cysteine codons (data not shown). The cassettes were created using the second-strand synthesis method of Reidhaar-Olson et al. 1991. The digested cassettes were inserted into pKSS.*smt3*-*mPheS*-FLAG and transformed into W3110 cells co-expressing Ulp1.

For purification of titin-FLAG constructs, his_6-*smt*3 was cloned into pET23b (Novagen) using *Nhe*I/*Not*I sites with the same *Age*I insertion as described above. Titin-FLAG constructs had N-terminal sequences as follows: X_5-MSHLA-*LIEVE*, where X_5 denotes the N-terminal candidate sequence followed by the first five residues of mPheS and the start of the titin domain in italics. Titin fragments were inserted using *Age*I/*Not*I sites.

The catalytic domain of SUMO protease Ulp1 (residues 403-621) was cloned into pET23b using *Nhe*I/*Bam*HI sites or purchased in purified form from Lifesensors Inc. (Malakhov et al. 2004). For expression *in vivo*, pKW221, encoding Ulp1 was constructed by inserting an *Avr*II/*Xho*I fragment containing codons 403-621 of Ulp1 fused to a C-terminal FLAG epitope under control of an arabinose-inducible promoter in the pACYC-based plasmid pJF122.

Media

YEG selection plates (Kast 1994) were made with the following modifications: 0.1% glycerol instead of glucose; 0.02% arabinose, and D,L-*para*-chloro-phenylalanine (Sigma).

Selections/Screens

Wild-type W3110 cells were transformed with the library of mPheS plasmids by electroporation and plated on YEG plates lacking Cl-Phe. Colonies were replica plated onto YEG plates with 0, 15, and 20 mM Cl-Phe and grown at 30°C overnight. To assess stability of mPheS *in vivo*, selected clones were grown at 30°C overnight in LB medium containing 0.05% arabinose. Saturated cultures were diluted 1:20 into LB medium with 0.5% arabinose and shaken at 37°C for 30 min. Spectinomycin was added to 200 μg/ml to halt protein synthesis, and 1 ml aliquots were removed at each timepoint. Cells were pelleted, resuspended in SDS buffer, and boiled prior to loading on a 10% Tris-glycine

gel. Western blotting was performed using M2 mouse α-FLAG (Sigma) at a 1:5000 dilution and an ECF detection kit (Amersham). Blots were imaged with a Typhoon 9400 imager (Amersham).

Pulse-chase analysis of mPheS degradation

Colonies were grown overnight in LB medium with appropriate antibiotics. Cultures were diluted 1:100 into 5 ml LB with antibiotics and 0.01% arabinose to induce low levels of SUMO protease. At OD_{595} between 0.6 and 0.9, cells were pelleted and washed once in M9 media supplemented with methionine-assay medium (Difco), plus trace metals and 0.4% glucose. Cells were resuspended in 250 μl of the same M9 medium. After 5 min shaking at 37°C, 100 μCi Expre^{35}S^{35}S label (Perkin Elmer) was added. After 5 min, chase solution (25 mg/ml each of L-Met and L-Cys) was added. At each timepoint, 50 μl was added to 50 μl lysis buffer (4% SDS, 125 mM Tris pH 6.8). Lysates were boiled for 5 min and added to 900 μl immunoprecipitation buffer (50 mM Tris pH 7.4, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100). Debris was pelleted at 13000 rpm at 4°C for 10 min, and supernatant was added to 0.4 μl α-FLAG agarose resin (Sigma) and shaken gently for 1 h at 4°C. Resin was pelleted and washed twice with 500 μl wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100), then dried, resuspended in 20 μl SDS buffer, and loaded onto 10% Tris-glycine gels. Dried gels were exposed to phosphorimager cassettes for 24 to 40 h and scanned on a Typhoon 9400 imager. Quantification was performed with ImageQuant 4.0.

Purification of titin substrates

His_6-SUMO-titin fusion proteins and Ulp1p were purified by Ni^{++}-NTA affinity chromatography according to Malakhov *et al.* (2004). After cleavage and removal of His_6-SUMO, titin protein was chromatographed over a Sephacryl-100 gel-filtration column (Amersham) in PBS buffer (20 mM sodium phosphate (pH 7.4), 150 mM NaCl).

Proteins were concentrated in Amicon Ultra centrifugal filters and their concentrations determined by absorbance at 280 nm. ^{35}S-labeled titin substrates were purified according to published protocols (Kenniston et al. 2003).

Degradation of tagged titin proteins by ClpAP in vitro

Degradation rates of ^{35}S-labeled titin variants were measured as a function of substrate concentration. ClpAP reaction conditions were as described in Weber-Ban et al. (1999) and degradation was initiated by addition of ATP regeneration mix containing 4 mM ATP, 50 mg/ml creatine kinase, and 5 mM creatine phosphate (Kim et al. 2000). Timepoints were removed at 30-second intervals over 3 minutes and quenched in 10% trichloroacetic acid (TCA). Degradation was quantified by measuring TCA-soluble peptides by scintillation counting as described. In the experiments presented in Figure 4, degradation rates were normalized to the total $ClpA_6$ concentration present in the reaction. $ClpA_6$ concentrations of 25, 50, or 100 nM were used depending on the substrate concentration (see Fig. 4 and legend) to ensure that substrate concentrations did not change more than 25% during the timecourse as a result of degradation. $ClpP_{14}$ was always used at twice the $ClpA_6$ concentration. Control experiments showed that degradation activity was linear with $ClpA_6$ concentration over this range. The dependence of the degradation rate of YLFVQ-titin (2 μM) on the ClpS:$ClpA_6$ ratio was also determined. Stimulation by ClpS saturated after six ClpS monomers per $ClpA_6$ under these conditions (data not shown; J. Hou, personal communication). All degradation reactions containing ClpS used a ratio of nine ClpS monomers per $ClpA_6$.

Acknowledgements

We are grateful to Jennifer Hou for providing strains and purified ClpS, and Chris Farrell and Sean Moore for development of the mPheS selection system. We also would like to thank Frank Solomon, Peter Chien, Saskia Neher, Jon Kenniston, Kathleen McGinness, and other members of the Baker and Sauer labs for helpful discussions. Supported by NIH grants GM-49224 and AI-16892. K.H.W. is supported by a National Science Foundation graduate fellowship. T.A.B. is an employee of the Howard Hughes Medical Institute.

Chapter Three

Tuning the Strength of a Bacterial N-end rule Degradation Signal

This work has been accepted for publication as Wang, K.H., Oakes, E.S.C., Sauer, R.T., and Baker, T.A. 2008. *J. Biol. Chem.* E.S.C. Oakes assisted in the surface Plasmon resonance binding assays and provided critical comments on the manuscript. R.T. Sauer and T.A. Baker assisted in preparation of the manuscript.

Abstract

The N-end rule is a degradation pathway conserved from bacteria to mammals that links a protein's stability *in vivo* to the identity of its N-terminal residue. In *E. coli*, the components of this pathway directly responsible for protein degradation are the ClpAP protease and its adaptor ClpS. We recently demonstrated that ClpAP is able to recognize N-end motifs in the absence of ClpS, although with significantly reduced substrate affinity. In this study, a systematic sequence analysis reveals new features of N-end rule degradation signals. To achieve specificity, recognition of an N-end motif by the protease-adaptor complex uses both the identity of the N-terminal residue and a free α-amino group. Acidic residues near the first residue decrease substrate affinity, demonstrating that the identity of adjacent residues can affect recognition even though significant flexibility is tolerated. However, shortening the distance between the N-end residue and the stably folded portion of a protein prevents degradation entirely, indicating that an N-end signal alone is not always sufficient for degradation. Together these results define the *in vitro* sequence and structural requirements of bacterial N-end signals.

Introduction

Regulated proteolysis is fundamental for cellular survival because it provides an irreversible control mechanism. For example, progression through the eukaryotic cell cycle requires timely turnover of cyclins to synchronize and order specific events such as DNA replication and chromosome segregation (Hershko 1997). Proteolysis also initiates the SOS response to DNA damage in bacteria via degradation of the transcriptional repressor LexA (Little 1983; Neher et al. 2003), thereby allowing expression of repair polymerases and checkpoint proteins (Walker et al. 1985). Defective protein turnover can initiate events as diverse as loss of competence in *Bacillus subtilis* (Turgay et al. 1998) and angiogenesis via hypoxia-inducible factor (Hif1) in mammals (Willam et al. 2002). The importance of proteolysis as a regulatory mechanism highlights the need to understand the mechanisms by which these proteases select the right substrates and avoid unintended protein destruction.

Energy-dependent proteases are composed of an oligomeric ATP-dependent unfolding enzyme and an enclosed proteolytic chamber (Sauer et al. 2004). The architecture of this chamber requires that substrates pass through an axial entry gate that is too narrow to allow entry of folded proteins. The unfoldase harnesses the energy of ATP hydrolysis to drive mechanical unfolding of protein substrates and to translocate the resultant denatured polypeptide into the proteolytic chamber for peptide-bond cleavage (Wang et al. 1997; Singh et al. 2000; Benaroudj et al. 2001). In *E. coli*, there are several proteolytic complexes; for example, the ClpX and ClpA ATPases unfold substrates and

translocate the polypeptide chains into the ClpP proteolytic chamber for degradation (Gottesman et al. 1998; Weber-Ban et al. 1999; Kim et al. 2000).

Known bacterial degradation signals vary in sequence complexity and in length from a few amino acids to roughly 10 residues (Flynn et al. 2003). Adaptor proteins also confer or enhance recognition by binding both the substrate and the unfoldase. For example, one region of the SspB adaptor binds to the ClpX unfoldase, and another domain recognizes a region of the ssrA degradation tag, facilitating tethering of ssrA-tagged substrates to ClpXP and the probability of productive engagement (Flynn et al. 2001; Wah et al. 2003). SspB and ClpX can bind the ssrA tag simultaneously, allowing SspB to hand off substrates to ClpX directly (Bolon et al. 2004). The sequence determinants for the ssrA•SspB and ssrA•ClpX interactions have been characterized both structurally and biochemically (Flynn et al. 2001; Levchenko et al. 2003; Song and Eck 2003). In contrast, the mechanism of adaptor-mediated delivery for substrates to ClpAP is not well understood.

ClpS is a multi-faceted adaptor, which enhances ClpAP turnover of N-end rule substrates but also prevents ClpAP from degrading other classes of substrates (Katayama et al. 1988; Dougan et al. 2002; Erbse et al. 2006; Wang et al. 2007; Hou et al. 2008). Because ClpS possesses both stimulatory and inhibitory activities, it can change the profile of ClpAP degradation dramatically. The evolutionarily conserved N-end rule relates the intracellular half-life of a protein to its N-terminal residue (Bachmair et al. 1986; Tobias et al. 1991; Varshavsky 1992). In bacteria, proteins beginning with

any of the three aromatic amino acids (Phe, Tyr, or Trp) or the aliphatic residue Leu are degraded by ClpAP with assistance from ClpS (Erbse et al. 2006). Side-chain hydrophobicity *per se* is not sufficient for N-end rule recognition, as Ile, Val, and Met do not target substrates for ClpAP degradation. Substrates with the same N-end rule residue but different adjacent sequences are also degraded with varying rates *in vivo*, indicating that residues beyond the N-terminus affect degradation in the bacterial N-end rule (Wang et al. 2007). It is known that ClpS binds directly to both ClpAP and N-end rule substrates to enhance protein turnover. ClpAP also shows weak affinity for N-end substrates in the absence of ClpS. These observations raise several questions about the mechanism of N-end rule substrate recognition by ClpA and ClpS. What is the molecular basis of the sequence signal that determines how efficiently an N-end motif is recognized? What are the individual contributions of ClpA and ClpS in degrading N-end motif substrates?

To address these questions, we mutagenized an N-end pentapeptide (YLFVQ) that efficiently targets substrates for ClpAP degradation (Wang et al. 2007) and assayed the effects on *in vitro* degradation of GFP-fusion proteins by ClpAP in the presence and absence of ClpS. We confirmed the importance of N-terminal Leu, Tyr, Trp, or Phe residues for robust ClpAPS degradation (Tobias et al. 1991). Competition experiments also established that modification of the α-amino group substantially diminished ClpAPS recognition. The N-end rule thus uses the combination of the N-terminal residue's side chain and the α-amino group as the principal recognition determinants of the degradation signal. The positive contributions of these two determinants are

antagonized by the presence of acidic residues adjacent to the motif, demonstrating that sequence adjacent to the N-terminal residue influences recognition by ClpAPS. Furthermore, N-end signals are not sufficient to promote degradation if the distance between the folded region of the protein and the N-terminal residue is too short, indicating that there is also a structural component to the N-end rule. Examination of individual contributions of ClpS and ClpAP revealed that ClpS bound poorly to acidic N-end motifs but well to short N-end motifs, whereas ClpAP degraded some acidic N-end substrates efficiently but could not degrade short N-end motifs. We conclude that both ClpS and ClpA are important in determining the efficacy of N-end substrate processing. These results dissect the bacterial N-end rule into components that are important for *in vitro* recognition and show how the presence of ClpS alters the sequence selectivity of ClpAP.

Results

Specific side chains at the N-terminal residue are critical for recognition

To allow the N-terminal sequence of a model substrate (GFP) to be modified without constraints imposed by translational initiation or post-translational processing, we constructed and purified variants as H_6-SUMO-X_7-GFP fusions, cleaved these proteins with SUMO protease (Malakhov et al. 2004), and repurified the X_7-GFP molecule to remove the protease and H_6-SUMO fragment. The strong N-end motif "YLFVQEL" was used as a reference X_7 sequence (Wang et al. 2007); the "EL" was encoded by a *Sac*-I restriction site to facilitate cloning. Variants with the first Tyr replaced by other N-end rule residues (Phe, Leu, Trp), by aliphatic side chains (Ile, Val), or by Thr were also constructed and purified. At low substrate concentrations where the rate of degradation by ClpAPS (ClpAP plus ClpS) was determined by the second-order rate constant (k_{cat}/K_M), only the N-end rule substrates were degraded efficiently (Fig. 3.1A; data not shown), consistent with the reported selectivity of the N-end rule (Tobias et al. 1991). Among good substrates, the variant with Phe at the N-terminus was degraded most rapidly, whereas the variant with Tyr was degraded at the slowest rate. This difference arose from a ~20 percent reduction in K_M but not V_{max} (not shown), suggesting modest differences in recognition of N-end residues by ClpAPS.

As another probe of the importance of the N-terminal residue, we assayed ClpAPS degradation of YLFVQEL-GFP in the presence of a large excess of peptide competitor consisting of a variable N-terminal residue followed by the first 21 residues of *E. coli* β-galactosidase. Efficient competition was observed when the N-end rule residues Tyr or

Leu were at the N terminus but not when Met or Arg occupied this position (Fig. 3.1B). Therefore, both direct degradation and competition assays can be applied to probe the sequence rules of the N-end signal.

The α-amino group is a recognition element of the N-end rule

The purified precursor H_6-SUMO-YLFVQEL-GFP protein was not degraded by ClpAPS (not shown), suggesting that the lack of a free N-terminal Tyr and/or the presence of "upstream" residues prevents recognition. To test the importance of a free α-amino group, we compared inhibition of ClpAPS degradation of YLFVQEL-GFP by the hexapeptide YLFVQR before and after blocking its N-terminus by treatment with acetic anhydride. The unmodified peptide was a good inhibitor, whereas competition by the acetylated peptide was reduced substantially but not eliminated (Fig. 3.1C). The latter result was not caused by incomplete acetylation, as MALDI-TOF mass spectrometry before and after chemical modification gave single species of the expected masses (Fig. 3.1D). These results indicate that the α-amino group is an important feature but is not an essential component of the N-end signal.

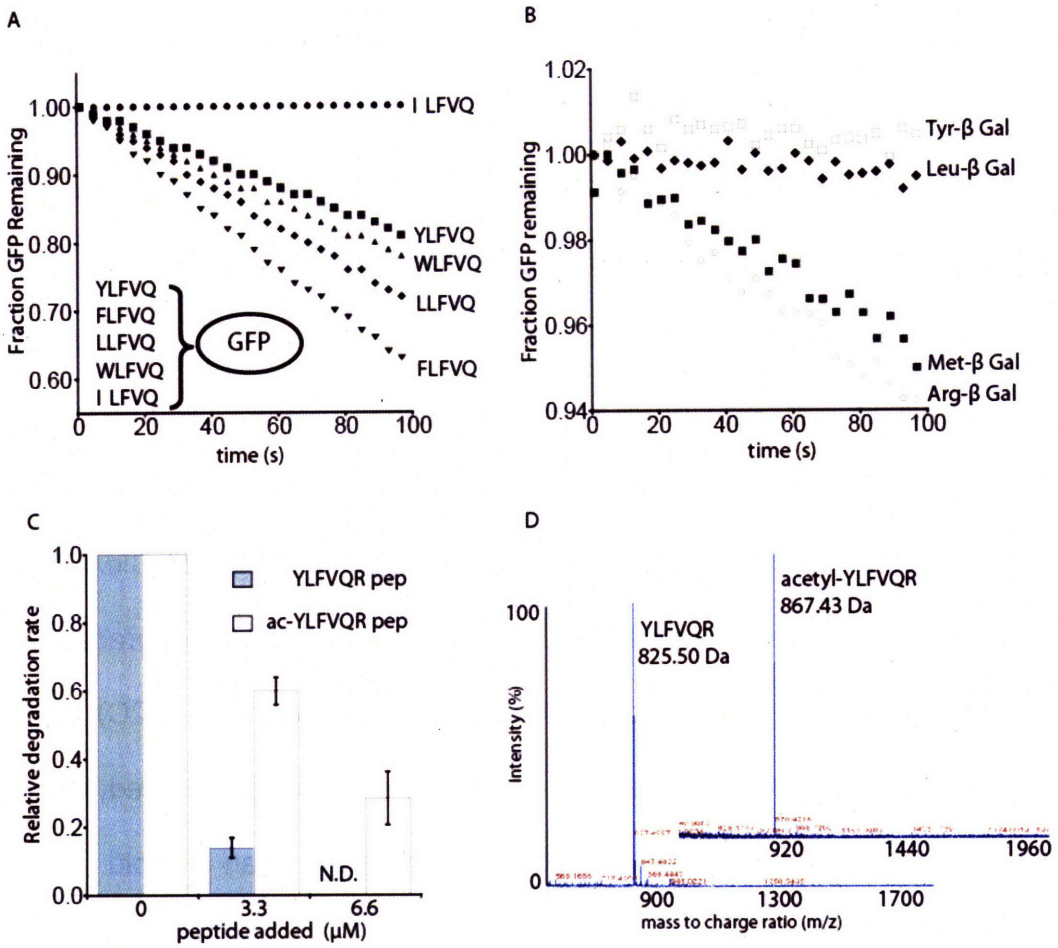


Fig. 3.1. The N-end rule depends on the identity of the first residue and the α-amino group. (A) N-end GFP constructs are degraded while the non-N-end ILFVQ-GFP is not. 50nM ClpAPS was incubated with 10nM GFP, and loss of GFP fluorescence was followed by excitation at 488 and detection at 511 nm. (B) Peptide competition of YLFVQ-GFP (500nM) degradation by 50 nM ClpAPS. β-Gal peptides (10 μM) with different N-terminal residues were added before initiation of degradation. (C) Peptide competition of YLFVQ-GFP (35nM) degradation by 50 nM ClpAPS. YLFVQR peptide with a free or acetylated α-amino group (ac-YLFVQR) was added before initiation of degradation, and rates were normalized to that of a reaction lacking peptide. N.D. indicates that the reaction using 6.6 μM YLFVQR peptide competitor was not performed. (D) MALDI-TOF spectra of unmodified versus acetylated YLFVQR peptide.

Acidic residues near the N-end residue weaken ClpAPS and ClpS binding

We previously found that substrates with the same N-end residue but different neighboring sequences were degraded with different K_M values, suggesting that residues beyond the N-terminus affect functional interactions with ClpAPS (Wang et al. 2007). To probe whether these effects are caused by packing or electrostatic interactions, we individually changed residues 2, 3, 4, and 5 of YLFVQEL-GFP to a basic residue (Arg), a small residue (Ala), or an acidic residue (Glu). When low concentrations of these substrates were tested for ClpAPS degradation, the Arg and Ala variants were degraded at rates similar to YLFVQEL-GFP (Fig. 3.2A, B), indicating that ClpAPS does not require specific side chains at positions 2-5 for efficient N-end degradation. By contrast, changing residue 2, 3, or 4 to Glu slowed degradation (Fig. 3.2C), with the largest effect observed when Glu was adjacent to the N-end residue. Indeed, replacing residue 2 with either Glu or Asp slowed degradation more than 10-fold, whereas changing this residue to Gln had only a small effect (Fig. 3.2D). Thus, the negative charge and not the shape of the position-2 side chain causes poor degradation by ClpAPS. A variant with residues 3-5 replaced by Glu (YL**EEE**EL-GFP) was degraded very slowly, suggesting that a net negative charge near the N-end residue is poorly tolerated by ClpAPS (Fig. 3.2D).

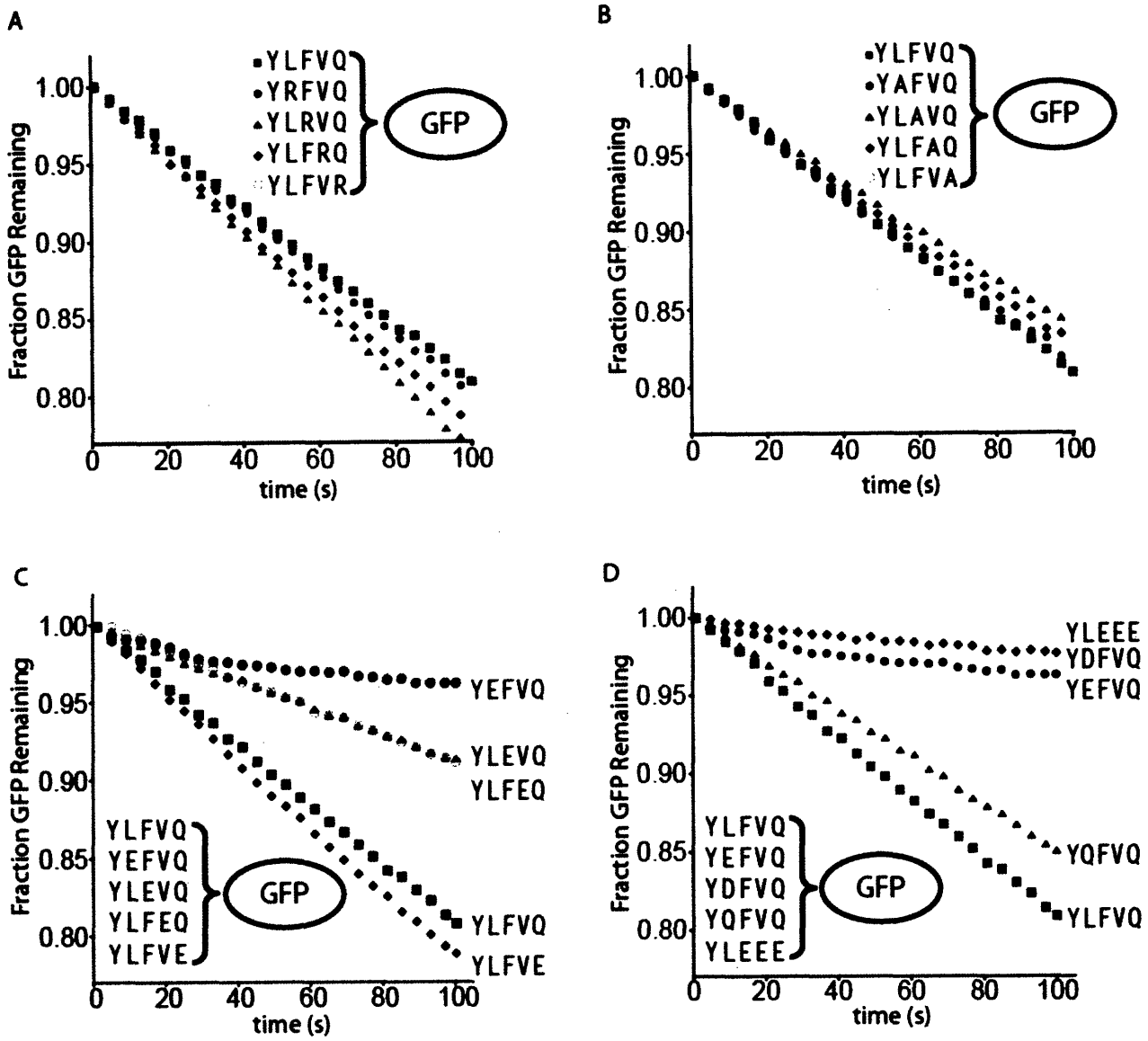


Fig 3.2. Acidic residues near the N-end residue slow degradation. Single Arg (A) or Ala (B) variants have negligible effects on GFP degradation. Initial rates of degradation for 10 nM GFP by 50 nM ClpAPS are shown. (C) Inhibition of N-end substrate degradation by an acidic residue is stronger when placed closer to the N-terminus. (D) An acidic residue but not Gln at position two or multiple Glu residues at positions three through five inhibit degradation.

To determine if the deleterious effects of acidic residues arose from poor substrate binding or slower turnover by ClpAPS, we determined steady-state kinetic parameters for the parental substrate YLFVQEL-GFP (K_M = 26 nM; V_{max} = 1.2 min^{-1}) and for YEFVQLE-GFP (K_M = 1400 nM; V_{max} = 1.4 min^{-1}) (Fig. 3.3). These results show that the principal effect of the $Leu^2 \rightarrow$Glu substitution is an approximate 50-fold weakening of apparent affinity of the substrate for ClpAPS. We conclude that acidic side chains at residues 2-4 of N-end degradation signals interfere with ClpAPS binding but not processing. The "N-end receptor sites" in ClpS and/or ClpA may have a negative electrostatic potential that interact unfavorably with negatively charged residues in the N-end signal.

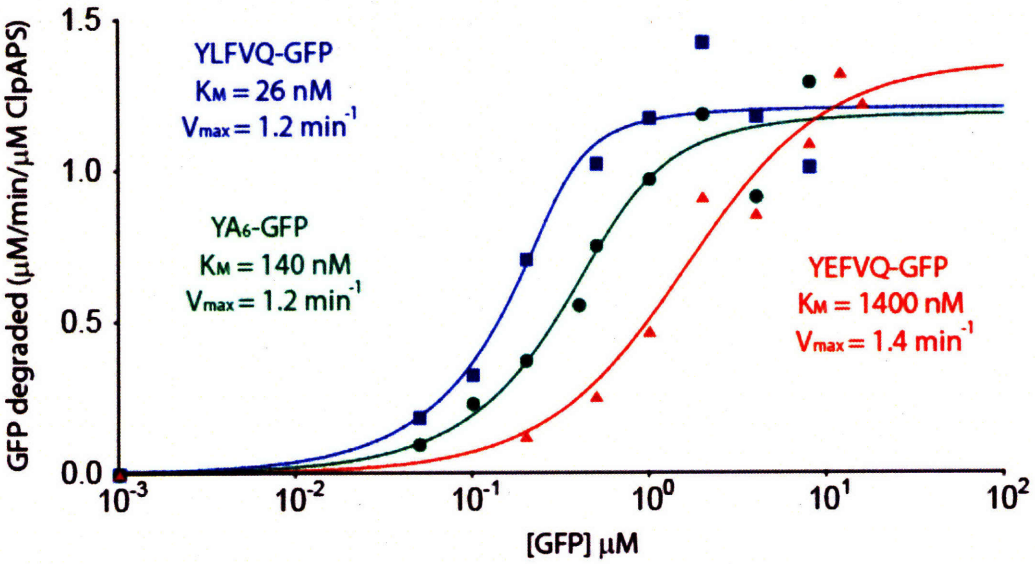


Fig. 3.3. The inhibition by acidic residues is caused by a reduction in affinity but not catalytic processing. Michaelis-Menten plot of initial degradation rates of YLFVQ-GFP at various GFP concentrations using 50 nM ClpAPS. The data represent the average of three experiments. YEFVQ-GFP is degraded with a 50-fold higher K_M but with a similar V_{max}. YA_6-GFP is degraded efficiently by ClpAPS but with a K_M value 6-fold higher than that of YLFVQ-GFP. Correlation coefficients (R^2) for all three curve fits were greater than 0.95.

To examine the relative affinity of ClpS for acidic N-end signals, surface plasmon resonance was used to monitor binding of immobilized ClpS to YEFVQEL-GFP and YLFVEEL-GFP. The YEFVQEL-GFP protein, which was degraded slowly by ClpAPS, also bound poorly to ClpS (Fig. 3.4A). By contrast, YLFVEEL-GFP has the same net charge but was degraded six-fold more rapidly by ClpAPS (Fig. 3.2C) and bound well to ClpS. These results show that acidic residues near the N-end residue influence ClpAPS degradation, at least in part, by weakening ClpS binding and also demonstrate that ClpS binding affinity is correlated with ClpAPS degradation activity of acidic N-end substrates.

Is this ClpS binding defect entirely responsible for the slow degradation of YEFVQEL-GFP by ClpAPS? ClpAP can degrade N-end substrates without ClpS, but with 10- to 70-fold weaker apparent affinity than ClpAPS depending on the sequence of the N-end signal (Wang et al. 2007). We found that ClpAP degraded both YEFVQEL-GFP and YLFVQEL-GFP at similar rates, indicating that ClpAP itself is not inhibited by an acidic residue at position 2 (Fig. 3.4B). In contrast, the presence of several acidic residues (YLEEEEL-GFP) near the N-end residue slowed substrate degradation by ClpAP. These results indicate that acidic N-end signals affect ClpAPS and ClpAP recognition differently.

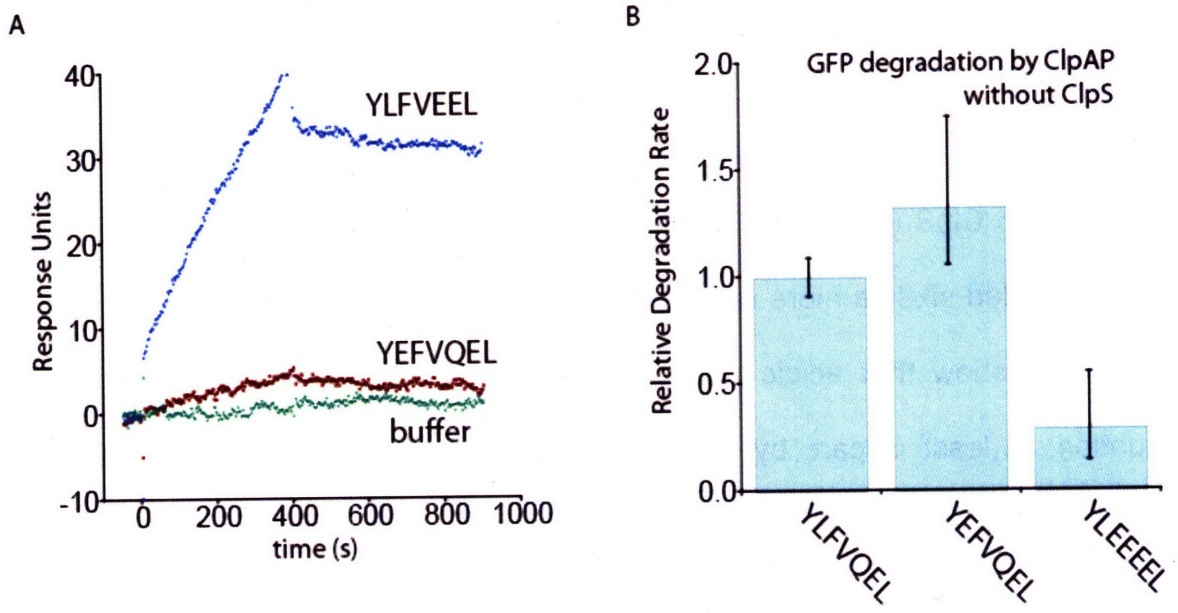


Fig. 3.4. A single Glu mutation in position 2 decreases affinity with ClpS but not ClpAP. (A) ClpS binds poorly to YEFVQEL-GFP but well to YLFVEEL-GFP by surface plasmon resonance. ClpS was immobilized to the chip surface and each GFP substrate was injected from time 0 to 400 s. The buffer curve is a comparison of responses of the chip surface from two separate buffer injections. (B) Degradation of acidic GFP constructs (8 μM) by 100 nM ClpAP without ClpS. Initial rates were normalized to that of YLFVQEL-GFP and the error bars show the error range of three independent reactions.

Length determinants of N-end signals

Erbse et al. (Erbse et al. 2006) found that ClpAPS degraded GFP with an N-end Phe followed by a 10-residue linker but not when the N-terminal Phe was placed adjacent to GFP. In our YLFVQEL-GFP construct, the N-terminal Tyr is seven residues from the alanine that begins the GFP sequence (YLFVQEL**A**SK; the lysine begins into the folded region of GFP). To address the role of linker length, we determined rates of ClpAPS degradation of constructs with six alanines between the N-terminal Tyr and the first residue of GFP (YA_6-GFP) and variants with the linker reduced by two (YA_4-GFP) or three residues (YA_3-GFP). The YA_6-GFP substrate was degraded by ClpAPS with a K_M of 140 nM, a value 6-fold higher than the K_M for YLFVQEL-GFP (Fig. 3.3). This result is consistent with a modest contribution of residues beyond the N-terminus to ClpAPS interactions. The YA_4-GFP substrate was degraded about 5-fold more slowly than YA_6-GFP (Fig. 3.5A), showing that linker length influences degradation. No degradation of YA_3-GFP was detected even at high substrate concentrations (Fig. 3.5B). Thus, GFP N-end tags that are too short pose a problem for ClpAPS.

To determine if this defect is due to the proximity of a folded domain adjacent to the YA_3 N-end signal, YA_3-GFP was acid-denatured prior to addition into a degradation reaction containing ClpAPS. Unfolded YA_3-GFP was degraded rapidly, whereas native YA_3-GFP was not turned over even using increased ClpAPS concentrations (Fig. 3.5C). This result indicates that N-end signals are not effective degradation motifs when located too close to the folded N-terminal region of the substrate. To establish whether

ClpS or ClpAP is responsible for this observation, experiments were designed to test the roles of both recognition modules.

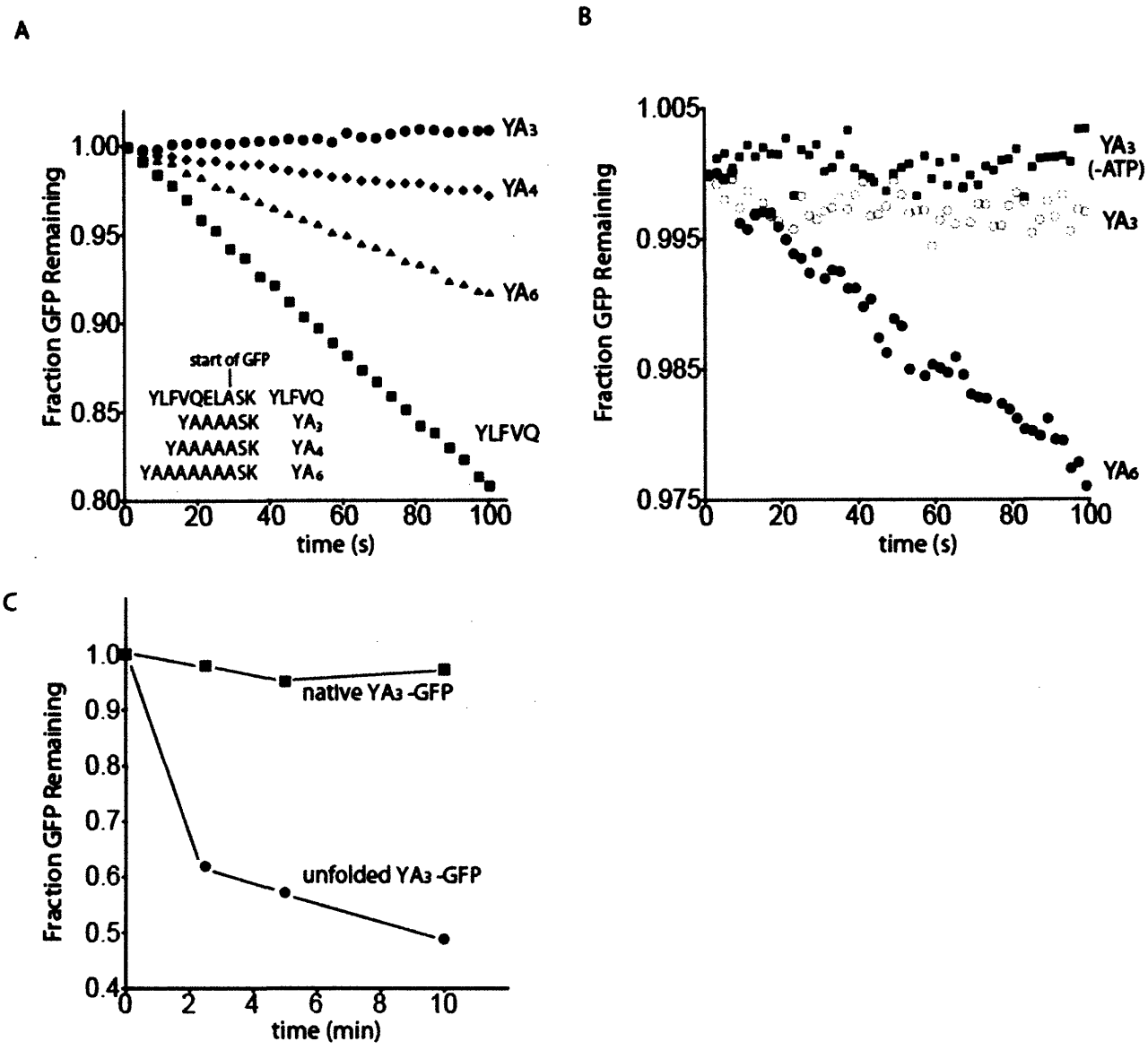


Fig. 3.5. ClpAPS requires a minimum tag length to degrade N-end motifs. (A) Degradation is reduced as the tag length is shortened. 10 nM GFP was incubated with 50 nM ClpAPS and loss of GFP fluorescence was followed over time. (B) YA_3-GFP is not degraded at 8 μM using 100nM ClpAPS. YA_6-GFP (8 μM) is shown for comparison. (C) Unfolded but not native YA_3-GFP is degraded rapidly by 800 nM ClpAPS. Protein was visualized by staining with Sypro Orange dye (Molecular Probes) and detecting fluorescence at 555 nm on a Typhoon imager.

Dissecting the individual contributions of ClpS and ClpAP to substrate recognition

In principle, ClpAPS might fail to degrade native YA_3-GFP either because ClpS does not bind this protein and/or because ClpA cannot accept this protein from ClpS or cannot unfold it after transfer. In surface plasmon resonance assays, immobilized ClpS bound YLFVEEL-GFP and YA_3-GFP to comparable extents but bound very poorly to ILFVQEL-GFP, a non N-end rule protein (Fig. 3.4A, 3.6A). To verify that ClpS is selective for N-end residues, fluorescinated peptides with N-terminal Phe, Tyr, Trp, and Ile were incubated with ClpS and fluorescence anisotropy was measured (Fig. 3.6B). All peptides except for the Ile variant produced an increase in anisotropy when ClpS was added, indicating that ClpS does not recognize an N-terminal Ile. Together, these data show that the inability of ClpAPS to degrade YA_3-GFP does not arise from a ClpS binding defect (Fig. 3.6C).

We determined rates of ClpAP degradation of YA_3-GFP, ILFVQEL-GFP, LLFVQEL-GFP, and YLFVQEL-GFP in the absence of ClpS (Fig. 3.6D). Under these conditions, the non N-end substrate (ILFVQEL-GFP) was degraded at a rate similar to the two good N-end rule substrates (LLFVQEL-GFP and YLFVQEL-GFP). However, ClpAP did not degrade the short-tag variant YA_3-GFP (Fig. 3.6D). Thus, N-end residues located too close to the folded region of GFP do not serve as degradation signals for ClpAP or for ClpAPS.

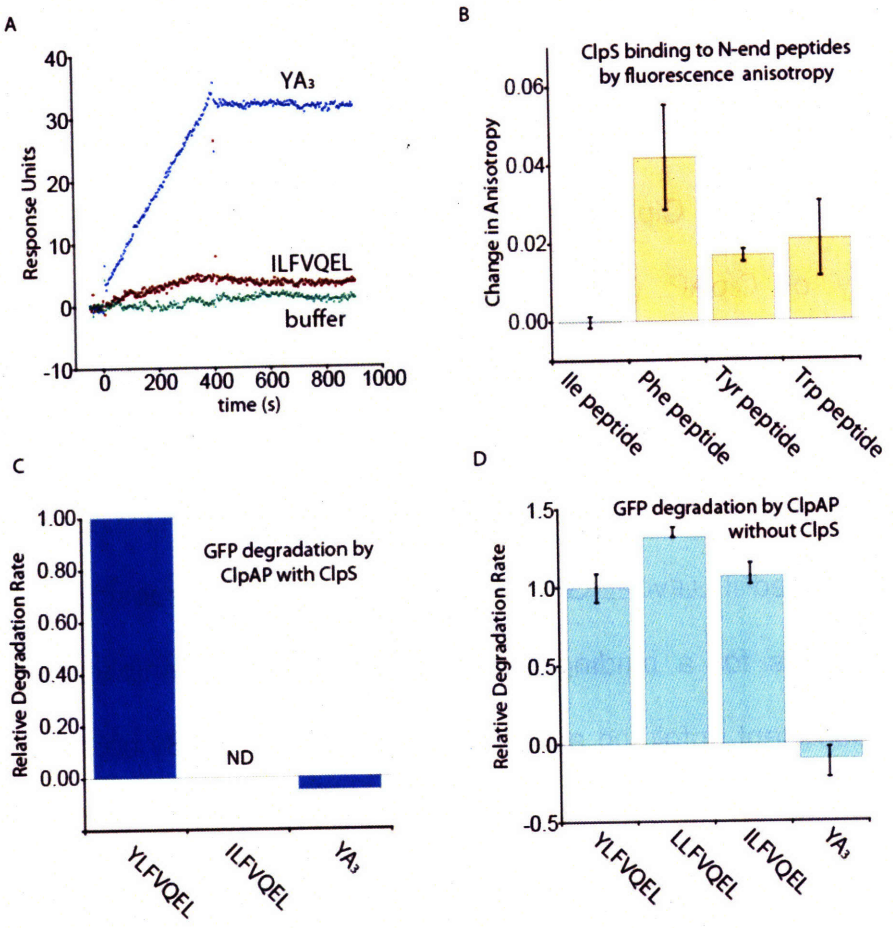


Fig. 3.6. ClpS and ClpA have distinct recognition specificities that contribute to the overall degradation efficiency of N-end substrates. (A) ClpS binds to the protein containing a short N-end signal YA_3-GFP but not the non N-end signal ILFVQEL-GFP by surface plasmon resonance. (B) Fluorescence anisotropy with labeled peptides shows that ClpS can distinguish Ile from the N-end residues Phe, Tyr, and Trp. Error bars show the error range of two independent trials. (C) Quantification of degradation of YLFVQEL-GFP, ILFVQEL-GFP, and YA_3-GFP by ClpAPS from Figs. 2A and 2C. Rates are normalized to that of YLFVQEL-GFP (ND indicates not detectable). (D) Degradation rates of LLFVQEL-GFP, ILFVQEL-GFP and YA_3-GFP by 100 nM ClpAP without ClpS. Initial rates were normalized to that of YLFVQEL-GFP and the error bars show the error range of three independent reactions.

The preceding experiments suggest that the defect in YA_3-GFP degradation arises after binding to ClpS. If this model is correct, then YA_3-GFP should act as a competitor and inhibit ClpAPS degradation of another N-end rule substrate by blocking the N-end recognition site on ClpS. In contrast, because ILFVQEL-GFP is not recognized by ClpS but is degraded by ClpAP, this substrate may compete by occupying the degradation activity of ClpAP (Fig. 3.7A). Indeed, ClpAPS degradation of the characterized N-end substrate ^{35}S-YLFVQMSHLA-titin (Wang et al. 2007) was inhibited by addition of YA_3-GFP, LLFVQEL-GFP, and ILFVQEL-GFP, but not by tag-less GFP (Fig. 3.7B). LLFVQEL-GFP was a much better competitor than YA_3-GFP or ILFVQEL-GFP. These results can be rationalized if LLFVQEL-GFP competes with the ^{35}S-substrate for binding ClpS but also competes for a binding site in ClpAP, thereby inhibiting both initial recognition and subsequent unfolding and degradation by ClpAP. By contrast, ILFVQEL-GFP is a weaker inhibitor because it only competes for ClpAP binding, and YA_3-GFP is a weaker inhibitor because it only competes for ClpS binding. Together with the results from Fig. 3.6, these data suggest that the efficiency of ClpAPS in degrading N-end rule substrates depends on recognition of the substrate by both ClpS and by ClpAP.

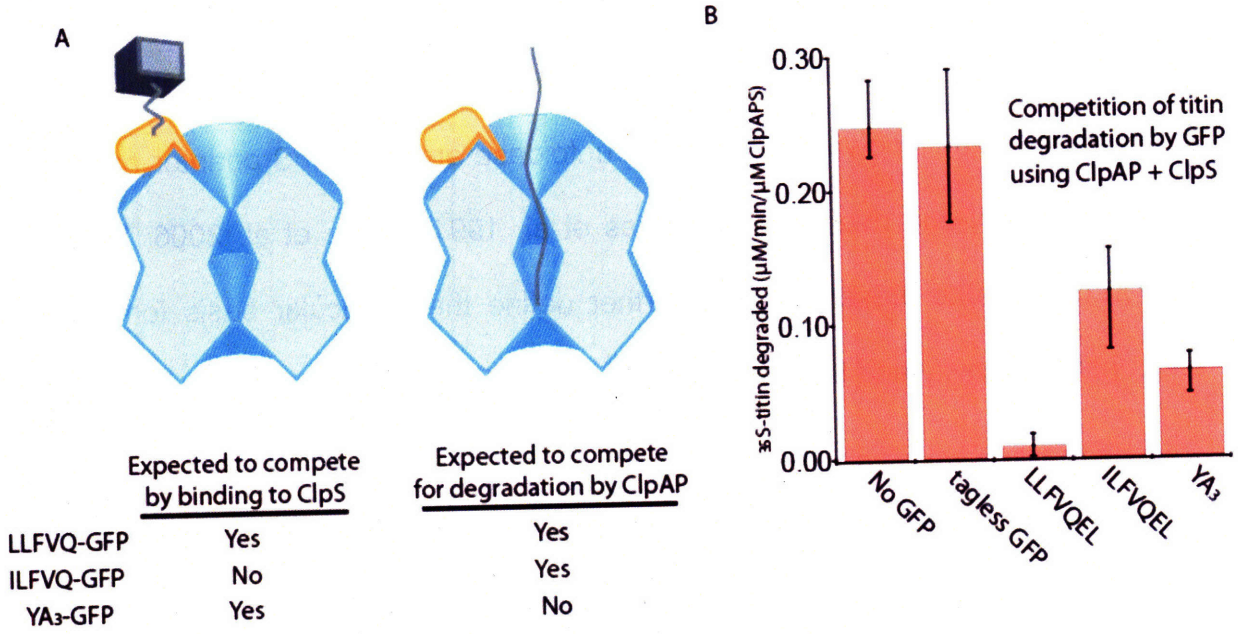


Fig. 3.7. Efficient competition is achieved by an N-end substrate that is recognized by ClpS and degraded by ClpAP. (A) Schematic predicting how different GFP substrates may compete with the N-end titin substrate used below. Binding to ClpS (orange) by GFP substrate blocks initial recognition of titin (gray), and degradation of GFP by ClpAP (light blue) prevents degradation of titin. (B) Inhibition of ^{35}S-labeled YLFVQMSHLA-titin degradation using 50nM ClpAPS and 10 μM GFP competitor substrate (40 μM in the case of tagless GFP). Degradation was quantified by counting the release of acid-soluble ^{35}S-peptides over time. Error bars for the bar graphs show the range of initial rates for three independent experiments.

Discussion

The original discovery of the bacterial N-end rule identified four N-terminal residues (Leu, Phe, Trp, and Tyr) that target β-galactosidase for degradation (Tobias et al. 1991). It is now known that the ClpA unfoldase and the ClpS adaptor participate in recognition of N-end rule signals (Tobias et al. 1991; Erbse et al. 2006; Wang et al. 2007). The results presented here further define the molecular basis for N-end rule sequence selectivity and the roles of ClpA and ClpS in recognition.

We confirmed that the expected N-end resdues mediated ClpAPS degradation of GFP variants with modest differences in efficiency in the order Phe > Leu > Trp > Tyr. By contrast, GFP with an N-terminal Ile showed no detectable ClpAPS degradation at low concentrations where N-end rule substrates were efficiently degraded. Thus, ClpAPS recognition is highly selective, discriminating between side chains as similar as Leu and Ile. In addition to the N-terminal side chain, we find that a free α-amino group contributes to but is not essential for ClpAPS binding. This finding is consistent with studies showing that blocking the N-terminus of an otherwise good N-end rule signal reduced ClpS binding on a peptide blot (Erbse et al. 2006). ClpS is required for high-affinity interactions with N-end rule substrates (Erbse et al. 2006; Wang et al. 2007), and we find that ClpS alone discriminates between substrates with good N-end rule residues and those with Ile at the N-terminus. Thus, ClpS enhances the degradation of N-end substrates by ClpAP by recognizing the α-amino group in combination with a Leu, Phe, Trp, or Tyr side chain at the N terminus (Fig. 3.8A).

Our results show that residues adjacent to the N-end residue influence the affinity of ClpAPS interactions. Specific side chains at these positions are not required. For example, changing the N-end signal of YLFVQEL-GFP to YAAAAAA-GFP increased the K_M for degradation only 6-fold, indicating that residues make small contributions to apparent affinity. Notably, however, an acidic residue at position 2 (YEFVQEL-GFP) increased K_M 50-fold; a variant with acidic residues at positions 3-5 (YLEEEEL-GFP) was also a very poor ClpAPS substrate. These effects are probably caused by repulsion between acidic residues in ClpAPS and those in these N-end signals, which is consistent with the slower *in vivo* degradation of substrates carrying acidic N-end sequences (Wang et al. 2007). Indeed, we found that ClpS alone bound YEFVQEL-GFP very poorly compared to YLFVEEL-GFP, and mutational studies suggest that Asp35 and Asp36 of ClpS form part of its binding site for N-end signals (Erbse et al. 2006). Negative electrostatic potential in this binding site would help bind the positively charged α-amino group of N-end signals. Moreover, some endogenous N-end signals contain a basic residue at position 2 because aminoacyl transferase adds Leu or Phe to bacterial proteins with an N-terminal Lys or Arg (Tobias et al. 1991). Hence, it seems likely that discrimination against acidic N-end sequences is a consequence of optimizing binding to N-end signals with an overall positive charge.

Importantly, our results and those of Erbse et al. (Erbse et al. 2006) demonstrate that proteins with N-end signals bind ClpS but are not necessarily ClpAPS substrates. Specifically, in our work ClpAPS and ClpAP did not degrade YA_3-GFP, even though ClpS bound this protein well. By contrast, ClpAPS degraded a variant with one extra

residue between the N-end Tyr and GFP (YA_4-GFP), although less rapidly than it degraded a substrate with a longer linker (YA_6-GFP). Apparently the distance between the N-end residue and the folded region of GFP must be sufficiently long to allow degradation, but this requirement is obviated when YA_3-GFP is unfolded. This length dependence could arise because steric restrictions prevent access of short GFP N-end tags to a binding site in the ClpA hexamer. Alternatively, such tags might be engaged by ClpA but be too short to allow a strong enough grip to allow unfolding.

Based on our results, we propose that N-terminal sequences have a wide range of abilities to target native proteins for ClpAPS degradation (Fig. 3.8B). At one extreme are short tags like YA_3, which do not target GFP for degradation, even though they have an N-end residue and bind ClpS well. Next are tags like ILFVQEL that do not have an authentic N-end residue or acidic N-end signals such as YEFVQEL that do not bind ClpS but can be engaged by sites in ClpAP. In the middle of the spectrum are signals with N-end residues that that have weaker affinities for ClpAPS because of the presence of negatively charged residues; both the number and positions of acidic residues appear to determine precise affinity. At the other extreme are strong N-end signals, such as YLFVQEL, that allow efficient ClpAPS degradation at nanomolar substrate concentrations.

Our results also propose several questions regarding the eukaryotic N-end rule, which recognizes the additional N-end residues Ile, Arg, and Lys. The N-end signal receptor Ubr1 is an E3 ubiquitin ligase that possesses a binding site for N-terminal Phe, Leu, Trp, Tyr, and Ile and a separate site for N-terminal Arg and Lys (Varshavsky 1996).

Interestingly, N-terminal Asp and Glu are recognized by the argininyl transferase Ate1p (Balzi et al. 1990), which conjugates an Arg residue to these N-termini. Is an N-terminal Arg residue recognized less efficiently when the second residue is acidic? If acidic residues in Ubr1 are important for docking the α-amino group and the Arg for this type of N-end signal, then electrostatic repulsion with acidic residues on the N-end signal may reduce binding affinity just as in the case of ClpS. Additionally, are shorter N-end sequences bound by Ubr1, and are these substrates ubiquitinated efficiently? Does the proteasome possess the same steric requirements for N-end signal length as ClpAP even though N-end substrates carry poly-ubiquitin chains as their proteasome localization determinants?

At present, it is not known how N-end substrates for ClpAPS are generated in the cell. Proteins with good N-end residues do not arise from translation and normal post-translational processing because the initiator formyl-Met of proteins with second residues Phe, Leu, Trp, or Tyr is not removed by methionine aminopeptidase (Flinta et al. 1986). The next challenge will be to isolate endogenous N-end substrates and to determine the extent and impact of sequence control in the N-end rule degradation.

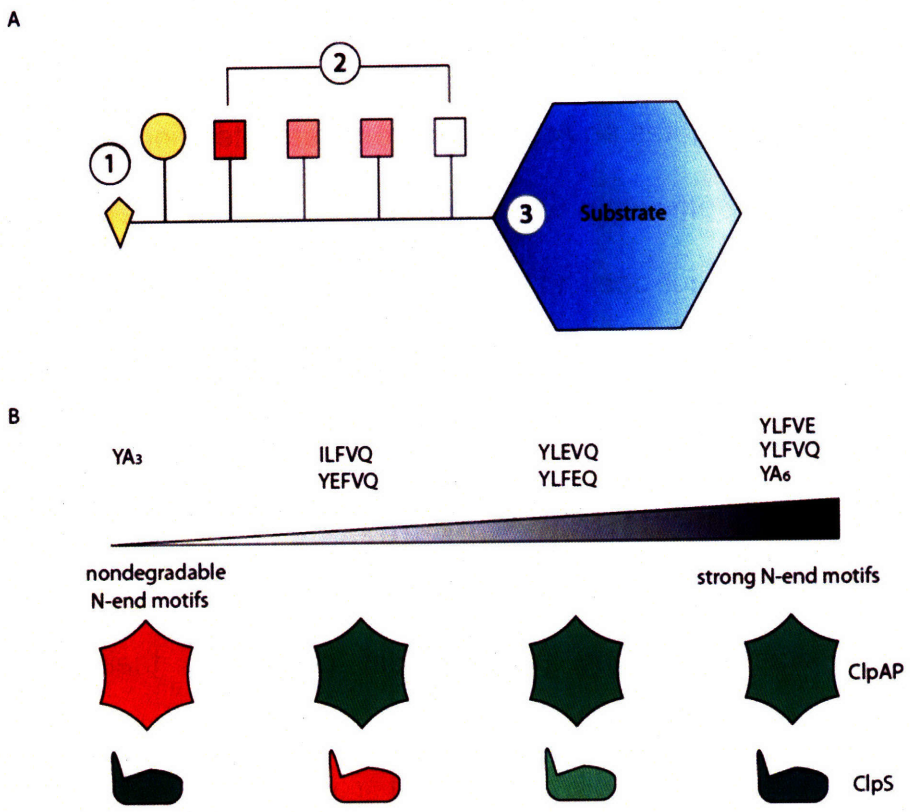


Fig. 3.8. (A) Components of an N-end signal. Recognition depends on the α-amino group and identity of the first residue (yellow diamond and circle, respectively; component 1) and is modulated by the neighboring amino acids (red and white squares, component 2). The residues immediately adjacent to the N-terminal residue modulate recognition more strongly than those further away. The structure of the folded region close to the N-end motif affects degradation of N-end substrates (blue, component 3). (B) A gradient model for N-end motif strength. Nondegradable N-end signals such as YA_3 are placed at the very left of the spectrum, and weak motifs somewhere in the middle depending on how efficiently they target substrates to ClpAPS for degradation. The relative affinities of ClpAP and ClpS to these degradation signals are represented by the red and green shapes below the spectrum. Red denotes negligible recognition, light green indicates moderate affinity, and dark green indicates a strong interaction.

Materials and Methods

Plasmids and Proteins

GFP variants (GFPuv with serine at position 65 changed to threonine) and Y-titin were cloned into a pET23b.smt3 vector using *AgeI* and *NotI* sites (Wang et al. 2007). The N-terminal sequences of GFP variants are shown in Fig. 5A; "ASK" initiates the GFP sequence.

For protein expression, substrates were subcloned into the IPTG-inducible vector pET23b.his$_6$-smt3 (pET23b from Novagen) and transformed into *E. coli* strain BL21 λDE3 (Malakhov et al. 2004). His$_6$-SUMO-GFP fusions were purified by Ni-NTA affinity chromatography as described (Malakhov et al. 2004) and were >85% pure. Most contaminants were His$_6$-SUMO or full-length His$_6$-SUMO-GFP. Some GFP variants were purified to >95% purity, using a low-substitution phenyl sepharose column (GE-Healthcare), but were degraded at the same rate as GFP proteins not processed with this second purification step. ClpA, ClpP, ClpS, and ^{35}S-YLFVQ-titin were purified as described (Katayama et al. 1988; Kim et al. 2000; Dougan et al. 2002; Kenniston et al. 2003; Wang et al. 2007). GFP (100 μM) was acid-denatured by adding hydrochloric acid to 25 mM for 5 minutes at room temperature (Hoskins et al. 2000).

GFP degradation assays

Loss of GFP fluorescence in degradation assays was monitored using a Photon Technology International fluorimeter (excitation 488 nm; emission 511 nm). ClpA$_6$ (50 nM), (ClpP-His$_6$)$_{14}$ (100 nM), ClpS (450 nM), and GFP substrate (10 nM) were premixed

as described at 30°C (Weber-Ban et al. 1999). For degradation reactions lacking ClpS, 100 nM $ClpA_6$ and 200 nM $ClpP_{14}$ were used. In figure 3, GFP concentrations from 50 nM to 16 μM were used. To initiate degradation, ATP (4 mM) was added at time 0. Initial changes in fluorescence were calculated from the linear portion of the kinetic trace, typically over the first 3 min, and converted to initial rates of GFP protein degradation using a linear standard curve relating fluorescence at 511nm to GFP concentrations. For determination of steady-state kinetic parameters in figure 3.3, the average initial rates from three independent experiments were plotted as a function of the total substrate concentration. Since [GFP substrate] was not always in excess of [ClpAPS], the data were fitted ($R^2 > 0.95$) by a non-linear least-squares algorithm to a quadratic version of the Michaelis-Menten equation:

$$V_{obs} = k_{cat} * \{(K_M + [ClpAPS]_o + [GFP]_o) - ((K_M + [ClpAPS]_o + [GFP])^2 - 4*([ClpAPS]_o*[GFP]_o))^{0.5}\} / (2*[ClpAPS]_o)$$

Degradation reactions of unfolded GFP were performed using $ClpA_6$ (800 nM), $(ClpP\text{-}His_6)_{14}$ (1.6 μM), ClpS (4.8 μM). ATP regeneration mixture (4 mM ATP, 50 mg/mL creatine kinase, and 5 mM creatine phosphate) was added prior to addition of unfolded GFP and the reaction was incubated at 30°C for 2 minutes. Unfolded YA_3-GFP (1.5 μM) was added at time 0 to initiate the reaction. At each timepoint, 10 μl of reaction mix was quenched by adding 2.5 μl of SDS loading buffer on ice. Samples were boiled and electrophoresed on a 15% Tris-glycine gel, which was stained with Sypro Orange (Molecular Probes) at a 1:5000 dilution in 7.5% acetic acid and scanned on a Typhoon

9400 imager (excitation 488 nm; detection 555 nm). Quantification was performed with ImageQuant 4.0, and intensities were normalized to the ClpP intensity in each lane. Three independent experiments were performed. A representative gel and quantification are shown in Fig. 3.5C.

Peptide competition assays

Peptide-competition assays were performed by assaying loss of YLFVQ-GFP fluorescence. β-galactosidase peptides were synthesized by the MIT Biopolymers facility and contained the first 21 residues of β-galactosidase fused to different N-terminal residues. These peptides were added to a final concentration of 50 μM in reactions containing 50/100/450 nM ClpA/P/S and 500 nM YLFVQ-GFP; degradation was started by adding ATP.

The YLFVQR peptide was acetylated by incubating 1 mM peptide in 10 mM Tris (pH 8.9) with 200 mM acetic anhydride overnight at room temperature. Acetyl-YLFVQR was purified by HPLC, lyophilized, and resuspended in H_2O. The addition of a single acetyl group was verified by MALDI-TOF mass spectrometry. Acetyl-YLFVQR or unmodified YLFVQR peptide were added to degradation reactions containing 50/100/450 nM ClpA/P/S, 35 nM of YLFVQ-GFP, and 3.3 or 6.6 μM peptide. The initial degradation rate in the absence of peptide was normalized to 1, and degradation rates in the presence of peptide competitor were determined relative to the initial rate and averaged (n=3).

Fluorescent labeling of peptides

Peptides with the sequence H_2N-XLFVQYH_6C (X represents different N-terminal residues) were synthesized using an Apex 396 solid-phase instrument, dissolved in 100 mM Tris (pH 7.5), and incubated with 5 μg/ml maleimide-fluorescein (Pierce) for 2 h at room temperature. Fluorescein-labeled peptides were purified by HPLC, lyophilized, and resuspended in water. Fluorescence anisotropy was measured at 30 °C (excitation 495 nm; emission 520 nm) using 1.4 μM fluorescinated peptide and 1.4 μM ClpS.

Protein competition assays

Samples containing 50/100/450 nM ClpA/P/S and 2 μM ^{35}S-YLFVQ-titin were premixed with GFP competitor substrate (10 μM except for untagged GFP [a gift of P. Chien], which was used at 40 μM). Degradation was initiated by addition of 4 mM ATP, and 10 μl aliquots were withdrawn every 30 s and quenched by addition of 10% trichloroacetic acid. Degradation rates were determined from the time-dependent accumulation of radiolabelled TCA-soluble peptides (Kim et al. 2000).

Surface Plasmon Resonance

ClpS binding experiments were performed using a Biacore 3000 instrument. ClpS was covalently bonded to a CM5 chip surface by amine coupling using the protocol supplied by the manufacturer. A 300 response unit (RU) surface of immobilized ClpS was used for the binding studies and another flow cell immobilized with 7000 RU of anti-ClpS antibody was used as a nonspecific binding control surface. GFP (440 nM) binding injections of 400 seconds were performed at a 30 μl/min flow rate in running buffer (20

mM HEPES (pH 7.5), 150 mM NaCl, 20 mM $MgCl_2$, and 0.005% P20 surfactant). Each GFP injection was preceded by an identical buffer injection whose composition matched that of the GFP solution. The GFP-ClpS interaction responses were double-referenced by subtracting the SPR signal from the GFP injection over the control flow cell as well as the signal from the buffer injection over the ClpS surface.

Acknowledgements

We wish to thank members of the Baker and Sauer labs for encouragement and helpful discussions; in particular, J. Hou for purified ClpS protein, P. Chien for untagged GFP protein, and I. Levchenko and G. Roman for synthesizing peptides. We are grateful to C. Köhrer and U. Rajbhandary for use of their Biacore 3000 instrument, and M. Pruteanu, A. Meyer, and S. Bell for critical reading of the manuscript. Supported by NIH grants GM-49224 and AI-16892. K.H.W. is supported by a National Science Foundation graduate fellowship. T.A.B. is an employee of the Howard Hughes Medical Institute.

Chapter Four

The Molecular Basis of N-end Rule Recognition

This chapter has been accepted for publication as "Wang, K.H., Roman, G., Grant R. A., Sauer, R.T., and Baker, T.A. (2008). *Mol Cell*". G. Roman assisted in the crystallization, degradation assays, and binding experiments. R.A. Grant led the structure determination and model building. R.T. Sauer prepared Figures 4.1B, C, D, and E and performed the sequence alignment in Figure 4.2A, and T.A. Baker prepared the submitted manuscript.

Abstract

The N-end rule targets specific proteins for destruction in prokaryotes and eukaryotes. Here, we report the crystal structure of a bacterial N-end rule adaptor, ClpS, bound to a peptide mimic of an N-end rule substrate. This structure reveals specific recognition of the peptide -amino group via hydrogen bonding and shows that the peptide's N-terminal tyrosine side chain is buried in a deep hydrophobic cleft in ClpS. The adaptor side chains that contact the peptide's N-terminal residue are highly conserved in orthologs and in E3 ubiquitin ligases that mediate eukaryotic N-end rule recognition. We show that mutation of critical ClpS contact residues abrogates substrate delivery and degradation, demonstrate that modification of the hydrophobic pocket results in altered N-end rule specificity, and discuss functional implications for the mechanism of substrate delivery.

Introduction

The N-end rule is a highly conserved mechanism that targets specific intracellular proteins for degradation. Certain N-terminal amino acids are recognized by the proteolytic machinery of the cell, resulting in the degradation of proteins bearing these residues (Bachmair et al. 1986). In bacteria, for example, aromatic and large hydrophobic residues (Tyr, Trp, Phe, and Leu) are primary N-end rule degradation signals (Tobias et al. 1991). In eukaryotes, these same N-terminal residues also serve as degradation cues, as do Ile and basic amino acids (Bachmair et al. 1986; Gonda et al. 1989). Importantly, these substrates are not generated by translation and standard N-terminal processing but by endoproteolytic cleavage, making N-end rule degradation especially suited for regulated proteolysis in diverse cellular processes (Mogk et al. 2007). The bacterial ClpS adaptor protein recognizes N-end rule substrates and delivers them to the AAA+ protease ClpAP for degradation (Bartel et al. 1990; Dougan et al. 2002; Tasaki et al. 2005; Erbse et al. 2006; Wang et al. 2007; Hou et al. 2008) (Fig. 4.1A). A specific family of E3 ubiquitin ligases recognizes and covalently modifies eukaryotic N-end rule substrates with ubiquitin, marking them for subsequent degradation by the proteasome (Bartel et al. 1990; Tasaki et al. 2005) (Fig. 4.1A). These E3 enzymes have a region of homology to ClpS (Lupas et al. 2003), suggesting that a common module mediates N-end rule recognition in organisms from bacteria to mammals.

Results and Discussion

The N-terminal tyrosine side chain fits into a hydrophobic pocket of ClpS

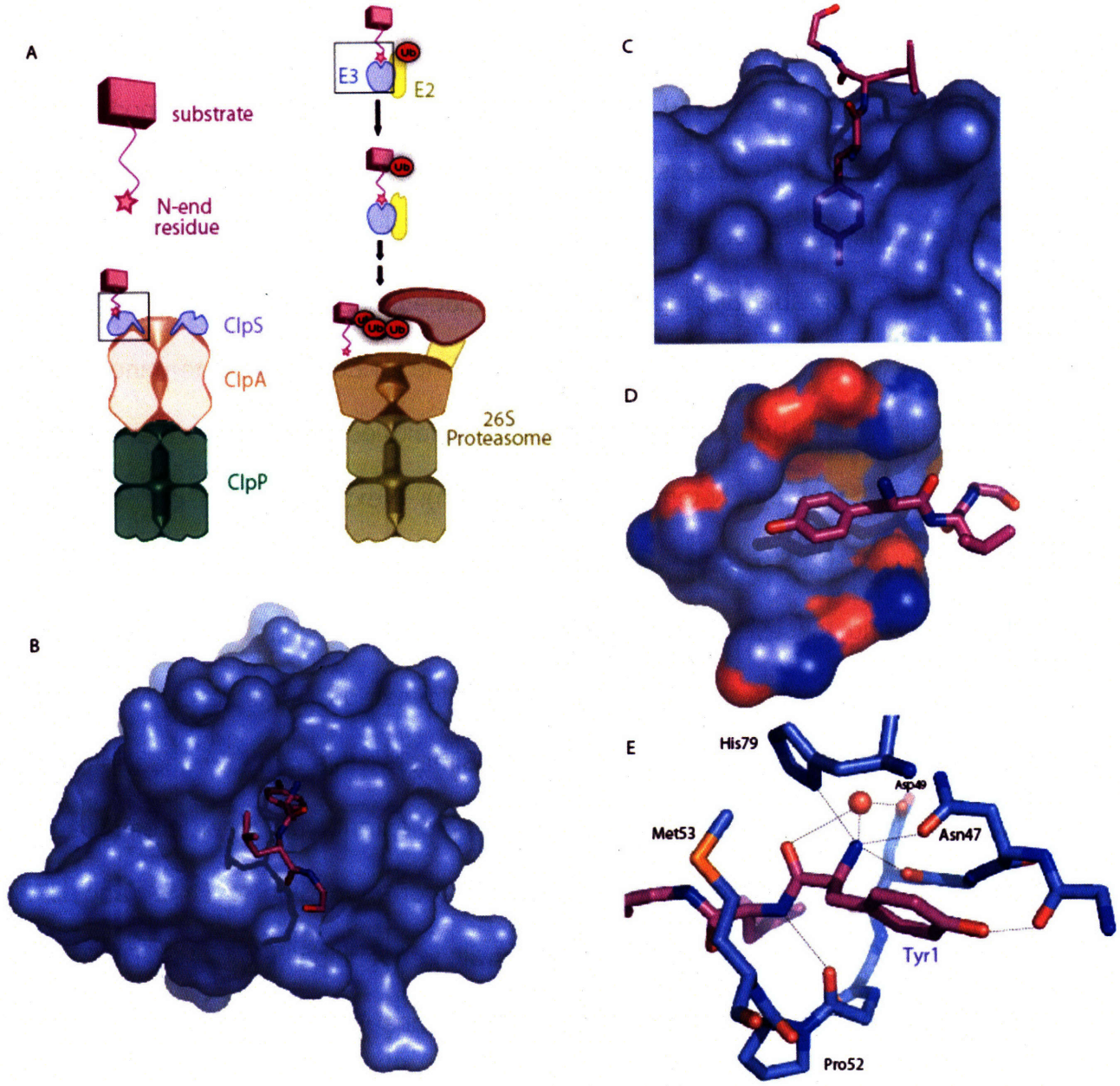
To determine how N-end rule substrates are recognized, we crystallized and solved the structure of the ClpS ortholog from *Caulobacter crescentus* bound to a peptide with an N-terminal tyrosine (pdb code 3D16; Fig. 4.1B, C, D). The ClpS fold in our cocrystal structure (2.0 Å resolution; R_{work} 17.7%; R_{free} 22.7%) was essentially the same as that of *Escherichia coli* ClpS (r.m.s.d. 0.5 Å for main-chain atoms) (Xia et al. 2004). This *E. coli* ClpS structure was determined without an N-end rule peptide and in complex with the N-terminal domain of ClpA (Zeth et al. 2002; Xia et al. 2004). Thus, neither binding to a peptide substrate nor to the ClpA N-domain, which occur on opposite faces of ClpS, appear to substantially change the adaptor's conformation.

The cocrystal structure reveals a simple mechanism by which ClpS recognizes bacterial N-end rule substrates. The peptide α-amino group, a unique chemical signature of the N-terminal residue, is hydrogen bonded to the side chains of *Caulobacter* ClpS residues Asn47 and His79 and to a water molecule that itself hydrogen bonds to the Asp49 side chain (Fig. 4.1E). Moreover, the tyrosine ring of the N-terminal peptide residue is buried in a deep hydrophobic specificity pocket on the surface of ClpS (Fig. 4.1C, D). At the bottom of this cavity, the tyrosine hydroxyl group donates a hydrogen bond to the main-chain oxygen of Leu46 (Fig. 4.1E). The specificity pocket of ClpS—which is formed by side-chain atoms from residues Ile45, Thr51, Met53, Val56, Met75, Val78, and Leu112—appears to be large enough to allow binding of a Trp side chain and could easily accommodate Phe and Leu side chains. Hence, we anticipate that ClpS will bind the N-terminal residues of all bacterial N-end rule substrates in a generally similar manner. ClpS also makes hydrogen bonds with the main chain of the second and third peptide residues (Fig. 4.1E), but specific side-chain contacts were not observed at these positions, and electron density for additional residues was absent. These results

are consistent with studies showing that residues past the N-terminus play only modest roles in determining the N-end rule binding affinity of ClpS (Wang et al. 2008).

Figure 4.1. N-end rule recognition. (**A**) N-end degron recognition. In bacteria, ClpS recognizes the substrate N-end rule signal (purple star) and directs substrate degradation by the ClpAP protease. In eukaryotes, a specific E3 ligase recognizes this signal and mediates addition of ubiquitin (Ub), which then leads to recognition and degradation by the proteasome. (**B, C**) In the cocrystal structure, the tyrosine side chain of the N-end rule peptide (purple, stick representation) sits in a deep hydrophobic pocket on ClpS (blue, surface representation). (**D**) Cutaway view of hydrophobic specificity pocket. The N-end rule peptide is shown in stick representation. *Caulobacter* ClpS residues 45, 51, 53, 56, 75, 78, 79, and 112 are shown in surface representation. Atom colors — oxygen (red); nitrogen (dark blue); sulfur (yellow); carbon (purple for peptide; slate blue for ClpS). (**E**) Key hydrogen-bond contacts (dotted lines) between ClpS and the N-end rule peptide. Same color scheme as panel D. Molecular graphics were prepared using PYMOL (DeLano, W.L. http://www.pymol.org)

1


A
substrate
N-end residue
ClpS
ClpA
ClpP
E3
E2
Ub
26S Proteasome
B
C
D
E
His79
Asp49
Met53
Asn47
Tyr1
Pro52

A hydrogen bonding network recognizes the α-amino group of the N-end signal

The functional relevance of our *Caulobacter* cocrystal structure is supported by strong evolutionary conservation of the ClpS side chains that make specific contacts with the N-end rule peptide (Fig. 4.2A). Nevertheless, we sought to test the importance of key ClpS residues directly. Because prior studies of the bacterial N-end rule have all been carried out with the *E. coli* protein, we first established that *C. crescentus* and *E. coli* ClpS have comparable binding specificities. Indeed, like *E. coli* ClpS, the *Caulobacter* adaptor bound tightly to peptides with Tyr, Trp, Phe, or Leu at the N-terminus (K_D 150 to 500 nM; Fig. 4.2B), but bound at least 20 to 50-fold less well to an otherwise identical peptide with Ile at the N-terminus (not shown). Next, based on our structure, we mutated *E. coli* ClpS residues predicted to contact the α-amino group of N-end rule substrates and assayed ClpS-mediated ClpAP degradation. The N34A mutation (corresponding to N47A in *Caulobacter* ClpS) eliminated detectable ClpAP degradation of a model N-end rule substrate (Fig. 4.2C). Similarly, this mutant variant of ClpS was unable to bind N-end rule peptides to a detectable extent even when present at 1.2 μM (not shown). The D36A and H66A mutations (corresponding to D49A and H79A in *Caulobacter* ClpS) also compromised substrate recognition (Fig. 4.2C, D). The D36A mutation caused a ~2-fold increase in the Michaelis constant (K_M) for substrate degradation, whereas the H66A mutation increased K_M about 5-fold and also lowered V_{max} substantially (Fig. 4.2D). We conclude that the structural interactions observed between ClpS and the α-amino group of the "substrate-mimic" N-end peptide play important roles in the recognition and delivery of N-end rule substrates for degradation.

2

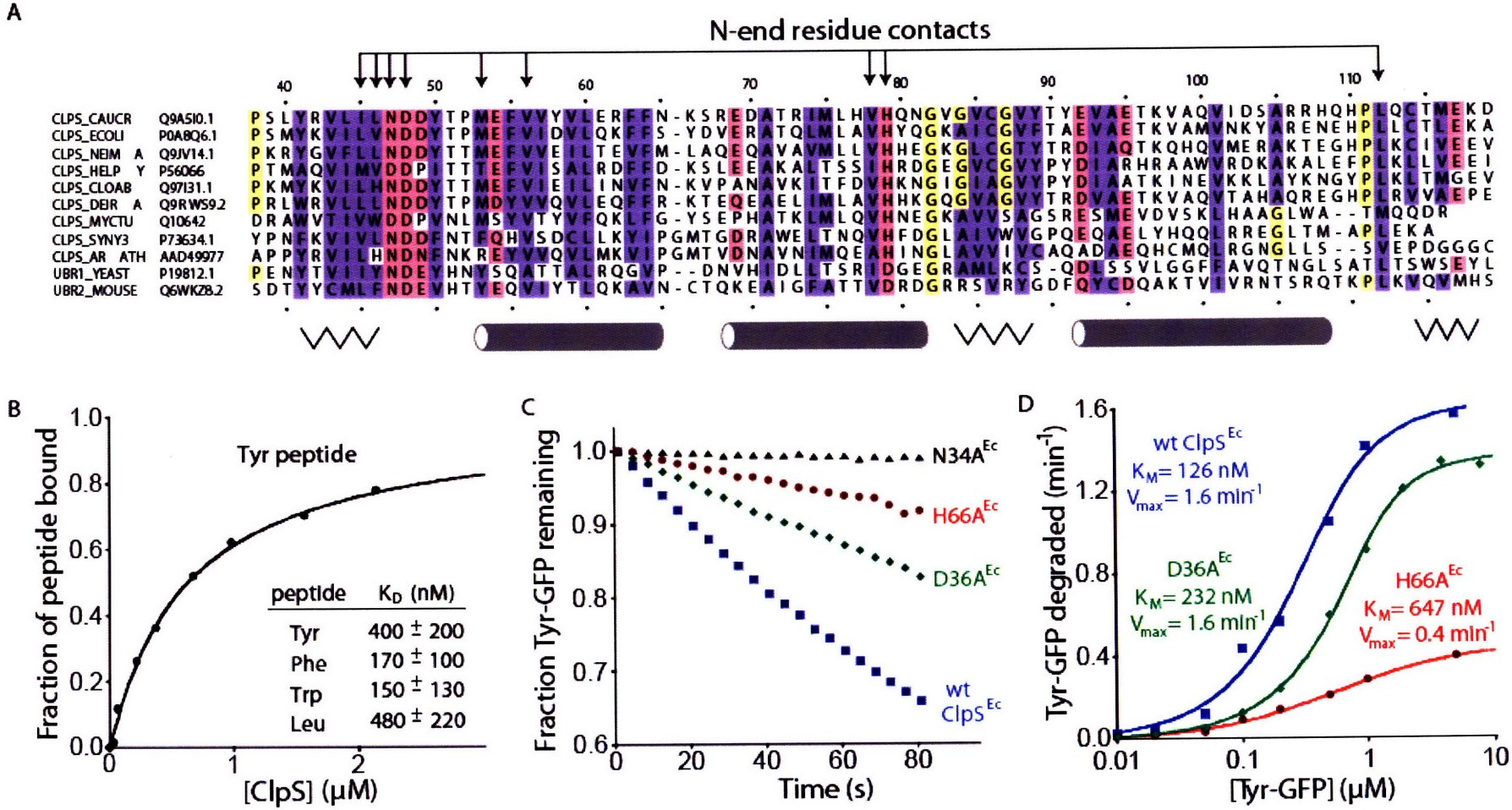

A
N-end residue contacts
40
50
60
70
80
90
100
110
CLPS_CAUCR Q9A5I0.1
CLPS_ECOLI P0A8Q6.1
CLPS_NEIM A Q9JV14.1
CLPS_HELP Y P56066
CLPS_CLOAB Q97I31.1
CLPS_DEIR A Q9RWS9.2
CLPS_MYCTU Q10642
CLPS_SYNY3 P73634.1
CLPS_AR ATH AAD49977
UBR1_YEAST P19812.1
UBR2_MOUSE Q6WKZ8.2
B
Tyr peptide
Fraction of peptide bound
[ClpS] (μM)
peptide
K_D (nM)
Tyr 400 ± 200
Phe 170 ± 100
Trp 150 ± 130
Leu 480 ± 220
C
Fraction Tyr-GFP remaining
Time (s)
N34A^Ec
H66A^Ec
D36A^Ec
wt ClpS^Ec
D
Tyr-GFP degraded (min^-1)
[Tyr-GFP] (μM)
wt ClpS^Ec
K_M= 126 nM
V_max= 1.6 min^-1
D36A^Ec
K_M= 232 nM
V_max= 1.6 min^-1
H66A^Ec
K_M= 647 nM
V_max= 0.4 min^-1

Figure 4.2. Conserved residues in ClpS are functionally important. (**A**) Sequence homology suggests conserved recognition of N-end rule substrates. The top eight ClpS sequences are bacterial. A plant ortholog (CLPS_ARATH) and two eukarytotic E3 ligases (UBR1_YEAST; UBR2_MOUSE) are also shown. Sequence numbering, secondary structure, and the positions of contacts with the N-end rule peptide are from the *C. crescentus* cocrystal structure. (**B**) *Caulobacter* ClpS binds N-end rule peptides beginning with Tyr, Phe, Trp, and Leu with sub-micromolar affinities. The curve shows ClpS binding to the Tyr peptide, assayed by changes in fluorescence anisotropy. (**C**) Mutations in *E. coli* ClpS ($ClpS^{Ec}$) that should alter direct or indirect contacts with the N-end rule α-amino group cause defects in ClpAP degradation of Tyr-GFP. The N34A, D36A, and H66A mutations in $ClpS^{Ec}$ correspond to N47A, D49A, and H79A in *Caulobacter* ClpS. Reactions contained $ClpS^{Ec}$ or variants (900 nM), ClpAP (100 nM), Tyr-GFP (100 nM), and ATP (4 mM). (**D**) Michaelis-Menten plots of Tyr-GFP degradation by ClpAP (100 nM) and $ClpS^{Ec}$ variants (900 nM). Representative plots for each variant and average K_M and V_{max} values are shown.

Met 53 is a specificity gatekeeper that excludes β-branched N-terminal residues

Lupus and Koretke (2003) originally reported sequence homology between ClpS and a subset of eukaryotic E3 ligases, and our structure establishes that the residues that form the ClpS N-end-rule binding pocket are highly conserved in these ligases (Fig. 4.2A). Thus, essential features of N-end rule substrate recognition are undoubtedly also preserved, although the ligases, but not ClpS, accept Ile as an N-terminal residue. Modeling indicated that this exclusion of Ile might arise from steric clashes between this β-branched side chain and the C_γ methylene group of Met53 in *Caulobacter* ClpS (Met40 in *E. coli*). To explore this possibility, we constructed and assayed the specificity of an *E. coli* M40A variant. The wild-type *E. coli* adaptor mediated efficient ClpAP degradation of green fluorescent protein with the N-terminal sequence Leu-Leu-Phe-Val-Gln-Glu-Leu (Leu-GFP), but showed little activity toward otherwise identical substrates with Val, Ile or Thr at the N terminus (Fig. 4.3A, C). By contrast, the M40A mutant delivered Ile-GFP to ClpAP as efficiently as Leu-GFP, and delivered Val-GFP better than either of these substrates (Fig. 4.3B, C). Importantly, the M40A variant retained the ability to recognize specific features of the N-terminal amino acid, efficiently delivering Val-GFP for ClpAP degradation but failing to deliver the isosteric Thr-GFP protein (Fig. 4.3C). Thus, this methionine side chain of ClpS serves as a specificity gatekeeper by excluding β-branched amino acids in bacterial N-end rule recognition. Furthermore, these results reveal that modest changes in the N-degron binding pocket could easily account for differences in recognition of hydrophobic N-end rule residues in prokaryotes and eukaryotes. Arg, Lys, and His are also N-end rule residues in eukaryotes but appear to be recognized by a different site in the E3 enzyme (Gonda et al. 1989).

3

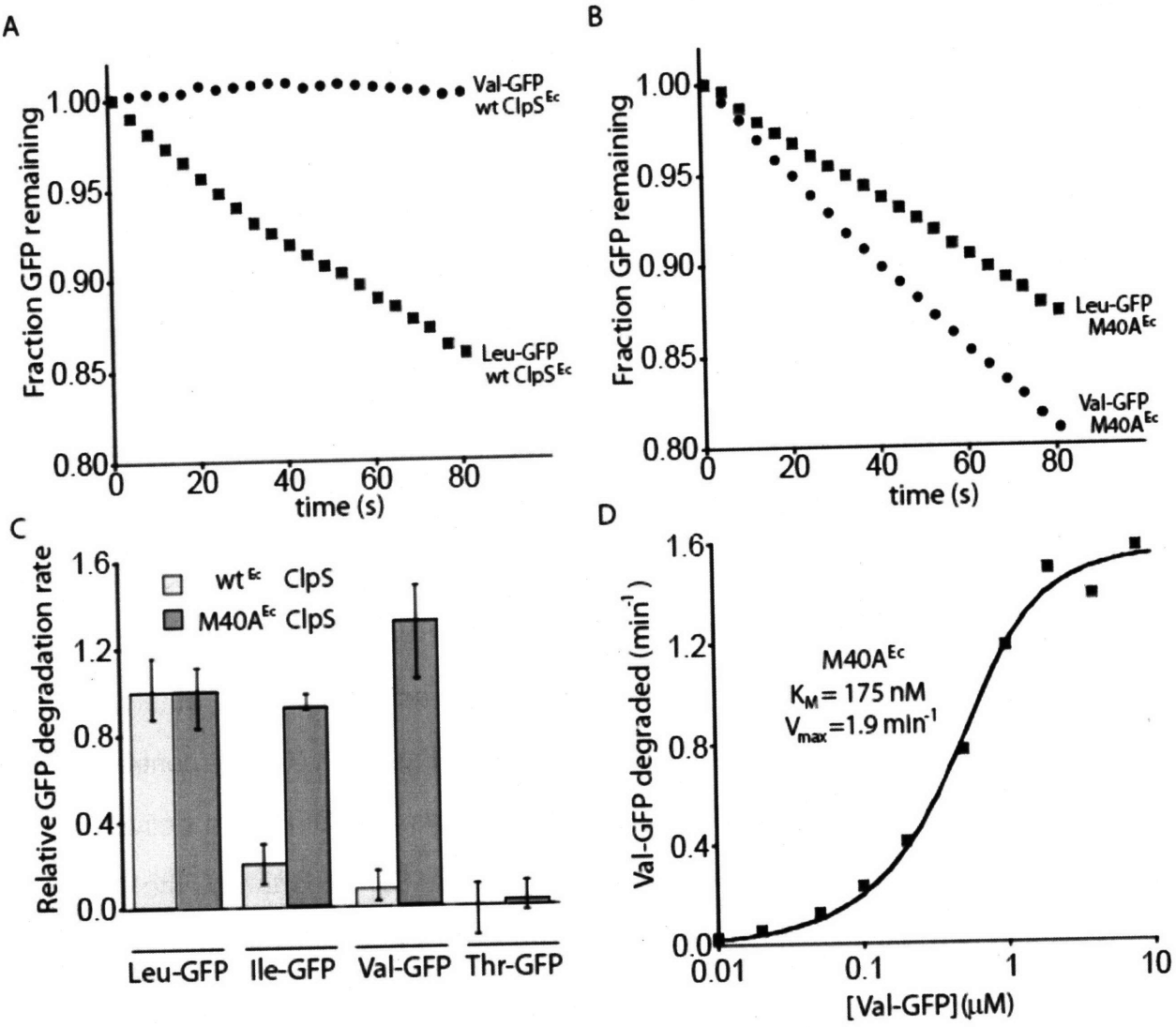

A
Fraction GFP remaining
1.00
0.95
0.90
0.85
0.80
0
20
40
60
80
time (s)
Val-GFP wt ClpS Ec
Leu-GFP wt ClpS Ec
B
Fraction GFP remaining
1.00
0.95
0.90
0.85
0.80
0
20
40
60
80
time (s)
Leu-GFP M40A Ec
Val-GFP M40A Ec
C
Relative GFP degradation rate
1.6
1.2
0.8
0.4
0.0
wt Ec ClpS
M40A Ec ClpS
Leu-GFP
Ile-GFP
Val-GFP
Thr-GFP
D
Val-GFP degraded (min-1)
1.6
1.2
0.8
0.4
0.0
M40A Ec
K M = 175 nM
V max =1.9 min-1
0.01
0.1
1
10
[Val-GFP] (μM)

Figure 4.3. Altered N-end rule degradation specificity. (**A**) Wild-type $ClpS^{Ec}$ delivered Leu-GFP but not Val-GFP for ClpAP degradation. (**B**) M40A $ClpS^{Ec}$ delivered Val-GFP for ClpAP degradation even better than Leu-GFP. (**C**) Rates of ClpS-mediated ClpAP degradation of GFP variants with different N-terminal residues were determined from experiments like those in panels A and B and normalized to the Leu-GFP rate. Ile-GFP and Val-GFP were delivered efficiently by M40A $ClpS^{Ec}$ but not by wild-type $ClpS^{Ec}$. Thr-GFP was delivered poorly by both adaptors. (**D**) Michaelis-Menten plot of Val-GFP degradation by ClpAP and $ClpS^{Ec}$ M40A. K_M and V_{max} are similar to the values for degradation of Tyr-GFP by ClpAP and wild-type $ClpS^{Ec}$ (Fig. 4.2D). In all panels, the ClpS concentration was 900 nM and the ClpAP concentration was 100 nM. Substrate concentrations in panels A-C were 500 nM.

Implications for transfer of N-end substrates from ClpS to ClpA for degradation

The ClpS•peptide cocrystal structure establishes the molecular basis of N-end rule recognition, which, together with ClpS binding to the N-terminal domain of ClpA is necessary but not sufficient for efficient ClpAP degradation (Hou et al. 2008; Wang et al. 2008). It is likely that release of the N-end rule side chain from its completely buried position in the ClpS pocket is a prerequisite for the next step of engagement and eventual transfer to the central pore of the ClpA enzyme. Possibly, residues in or near the ClpA pore facilitate hand-off by interacting with ClpS residues close to the binding pocket. For example, such interactions could reposition the ClpS side chains that contact the α-amino group of the substrate, destabilizing adaptor-substrate binding in a manner analogous to the N34A and H66A mutations in *E. coli* ClpS. Interestingly, mutations in two residues of *E. coli* ClpS (Y37A and E41A), which are near the binding pocket but do not contact the substrate, decrease the efficiency of substrate delivery (Erbse et al. 2006). Hence, we suggest that these adaptor residues facilitate a process of active substrate transfer from ClpS to the ClpA pore. The structure reported here should aid in the design of experiments to probe the mechanism of downstream delivery to the ClpAP protease.

Materials and methods

Residues 20-119 of *C. crescentus* ClpS were fused to the C-terminus of His_6-SUMO-Tyr-Arg-Gly using a pET23b vector (Novagen). Residues 1-19 were not included in this construct because the corresponding residues in *E. coli* ClpS are largely unstructured (Zeth et al. 2002; Xia et al. 2004). Following fusion-protein purification by Ni^{++}-NTA chromatography (Qiagen), cleavage with SUMO protease resulted in an insoluble ClpS fragment. However, cleavage with thrombin (Novagen) produced a soluble fragment, $ClpS^{35-119}$, which was purified by repassage through Ni^{++}-NTA, passage through a Mono-Q column (GE Healthsciences), and gel filtration on Superdex-75 (GE Healthsciences). Full-length variants of *E. coli* ClpS were constructed by PCR mutagenesis and purified as described (Dougan et al. 2002).

Crystals in space group $P2_1$ were obtained after 1 week at 20 °C in hanging drops containing 2 µl of protein solution (8 mg/ml *C. crescentus* $ClpS^{35-119}$, 2 mM of peptide Tyr-Leu-Phe-Val-Gln-Arg-Asp-Ser-Lys-Glu, 10 mM HEPES (pH 7.5), 200 mM KCl, and 1 mM DTT) mixed with 1 µl of reservoir solution (0.1 M bis-Tris (pH 5.5), 0.2 M $MgCl_2$, and 19% PEG 3350). Crystals were frozen without additional cryoprotection. X-ray diffraction data were collected on a Rigaku MicroMax007-HF rotating anode source equipped with Varimax-HR mirrors, an RAXIS-IV detector, and an Oxford cryo-system, and were processed using HKL-2000 (Otwinowski et al. 1997). Initial phases were obtained by molecular replacement using *E. coli* ClpS (1R60 chain C) as a search model in PHASER (Storoni et al. 2004). The final structure was obtained by iterative model building using COOT (Emsley et al. 2004) and refinement using PHENIX (Adams et al. 2002).

E. coli ClpA (Maurizi et al. 1994), *E. coli* ClpP-His_6 (Kim et al. 2000), and GFP substrates (Wang et al. 2008) with the N-terminal sequence Xxx-Leu-Phe-Val-Gln-Glu-Leu (where Xxx is a variable position) were purified and degradation experiments were performed as described (Wang et al. 2008). For binding assayed by fluorescence anisotropy, synthetic peptides (Xxx-Leu-Phe-Val-Gln-Tyr-His_6-Cys) were labeled by modification with fluorescein maleimide (Wang et al. 2008). K_D values were averages (± SD) of three independent experiments; K_M values were averages of two experiments.

Acknowledgements

We wish to thank James Partridge, Steve Brohawn, Thomas Schwartz, Steve Glynn, Igor Levchenko, and Peter Chien for advice on crystallization. We appreciate the encouragement from all members of the Baker and Sauer lab. This work was funded by NIH grants GM-49224 and AI-16892. K.H.W. is supported by a National Science Foundation graduate fellowship. T.A.B. is an employee of the Howard Hughes Medical Institute.

Chapter Five

Discussion: Beginning of the N-end?

Protein turnover is an effective method of downregulating a cellular activity because it is irreversible. However, because degradation by ATP-dependent proteases often is complete and rapid, substrate selection must be tightly regulated by the recognition of specific degradation signals. Below I discuss how the N-end signal can serve as an adjustable degradation signal in the cell. Additionally, I speculate on strategies to isolate endogenous N-end substrates and present several models for N-end substrate delivery from the ClpS adaptor to the ClpAP protease.

Significance of N-end sequence rules and signal strength

Modulating the rate of N-end degradation in vivo

The ability of ClpAPS to recognize N-end signals with a wide range of affinities allows the cell to tune N-end degradation depending on the need for a particular substrate to be turned over rapidly or completely. For example, N-end degradation can be slowed by the presence of acidic residues in the N-end signal. This effect was observed in the 5-fold slower degradation of WECVE-mPheS versus WFCWS-mPheS *in vivo*, suggesting the possibility of modulating the turnover of endogenous N-end substrates through sequence context. Substrates that need to be eliminated completely should possess N-end signals that are recognized efficiently, meaning that they should lack acidic residues and be of a sufficient length to be degraded by ClpAPS.

The suppression of V_{max} in the presence of ClpS (chapter 3) could have cellular implications if there are large amounts of deleterious N-end proteins that need to be cleared effectively. In this situation, ClpS slows N-end degradation relative to the degradation rate by ClpAP without adaptor because ClpAP's low affinity for N-end

signals is overcome by substrate abundance. However, even abundant proteins such as EF-Tu are present at only concentrations of ~30 μM in the cell (assuming a cell volume of 10^{-15} L) (Neidhardt and Umbarger, 1996; Ishihama et al. 2008).Since the K_M for degradation in the absence of ClpS is 30 to 90 μM based on the substrates in chapter 2, the Vmax suppression effect by ClpS is not likely to be a factor *in vivo*. In addition, ClpS stimulates degradation of N-end proteins with K_M values ranging from 26 nM to 3 μM and therefore should prevent the accumulation of large amounts of any N-end protein.

Searching for endogenous N-end substrates

Enrichment of N-end signals in a selection for N-terminal degradation sequences

The results of the selection system described in chapter 2 highlight several features of the N-end rule. Firstly, N-end degradation signals depend primarily on the identity of the N-terminal residue, whereas substantial sequence variation is tolerated after the first residue. N-end signals are efficiently recognized *in vivo*, and coupled with the low sequence complexity required, this explains the enrichment of N-end sequences in the mPheS selection. In other words, N-end signals are abundantly encoded in my selection and are efficient degradation signals, leading to the unexpectedly high yield of positive clones.

Endogenous N-end signals are not likely revealed by methionine aminopeptidases

How might the N-end signals isolated from the mPheS selection lead to the identification of endogenous N-end signals? Obviously Phe, Tyr, Trp, and Leu residues

can be found throughout most protein sequences, but exposure of these residues at the N-terminus requires proteolytic processing. One general mechanism of processing is the removal of the initiator Met residue by amino peptidases (Flinta et al. 1986). These peptidases cleave efficiently after the first Met if the radius of gyration of the second residue is small (Dalboge et al. 1990). Because most N-end residues in both bacteria and eukaryotes are large, it is unlikely that a significant pathway of N-end signal generation occurs by processing of the N-terminal Met.

Proteolytic processing reveals latent N-end signals

Eukaryotic N-end substrates are often isolated with prior knowledge of their cognate proteases. Many of the examples highlighted in chapter I was discovered after observing that the instability of the protein depended on proteolytic processing. Therefore, any extrapolation of the conclusions from the selection assay and the sequence rules in chapter 3 to the *in vivo* N-end rule should begin with a search for endopeptidases. A survey of endopeptidases on the Expert Protein Analysis System (Expasy) identifies a select few enzymes with cleavage specificities that are flexible at the residue position directly following the cleavage site, making these enzymes good candidates for generating N-end signals. For example, caspases cleave after specific four-residue sequences as long as the following residue is not charged or Pro. As mentioned in chapter I, the instability of DIAP depends on caspase activity (Ditzel et al. 2003).

An intriguing candidate N-end protease is pepsin, an enzyme secreted by chief cells in the stomach lining. Pepsin specifically cleaves in front of the residues Phe, Tyr,

Trp, and Leu. The argument against the role of pepsin as an N-end endopeptidase is that it acts on extracellular substrates and only under extremely acidic conditions (below pH 2). Substrates of pepsin are therefore likely to be degraded by other digestive enzymes rather than ubiquitinated and targeted to the proteasome, which is intracellular.

Bacterial peptidases that do not rely on the identity of the residue following the cleavage site are candidates for the generation of N-end substrates. For example, staphylococcus peptidase (*S. aureus*) and clostripain (*C. histolyticum*) cleave after an acidic residue regardless of the identity of the following position, but it is unclear if either of these bacterial species possesses ClpS or ClpA. Thermolysin (*B. thermoproteolyticus*) is an interesting candidate peptidase because it cleaves before Ala, **Phe**, Ile, **Leu**, Val, and Met (N-end residues in **bold**) unless the previous residue is acidic. However, thermolysin is like pepsin in that both are destined for secretion.

A strategy to isolate endogenous bacterial N-end substrates

Based on the work in chapters 3 and 4 demonstrating that ClpS binds specifically to N-terminal Phe, Tyr, Trp, and Leu residues, ClpS-interacting proteins may be *bona fide* N-end substrates for ClpAP. One method of identifying endogenous N-end substrates is therefore to purify ClpS in complex with bound proteins. In collaboration with J. Hou (Baker lab), ClpS was His_6-tagged at the C-terminus and expressed in Δ*clpS* Δ*clpA* cells. As a control, ClpS (D35A / D36A) was also expressed; this mutant binds poorly to N-end peptides *in vitro* and does not promote degradation of N-end

substates by ClpAP (data not shown; Erbse et al. 2006). The comparison of interacting proteins in a sample using ClpS-His_6 (D35A / D36A) can therefore be subtracted from the list of proteins pulled down by ClpS-His_6 to obtain a group of potential N-end substrates.

To validate candidate proteins as N-end substrates, one needs to identify the mature N-terminal sequence *in vivo*. This can be performed by overexpressing each candidate with a C-terminal epitope tag and affinity-purifying the candidate protein from lysate. The purified protein should then be N-terminally sequenced by Edman degradation. To enrich for ClpS-interacting isoforms of the candidate protein, ClpS-His_6 can instead be used to purify candidate proteins from the lysate. Another method of validation would be to perform a far western blot using ClpS to probe a membrane containing lysate from cells overexpressing the candidate proteins. ClpS should bind specifically to proteins bearing N-end residues, allowing for the identification of proteolytically processed isoforms of the candidate protein that bear N-end signals.

Delivery of N-end substrates from ClpS to ClpAP

From the work in the preceding chapters it is clear that ClpS and the ClpA hexamer ($ClpA_6$) both recognize N-end signals, but the mode of substrate handoff remains unknown. In addition, ClpS and $ClpA_6$ may both use the N-terminal residue as a binding determinant. If this is indeed true, how can $ClpA_6$ access the N-end residue when it is buried in the hydrophobic pocket of ClpS?

Role for the N-terminal tail of ClpS

Deletion of the first seventeen residues of ClpS prevents delivery of N-end substrate to $ClpA_6$ (Hou et al. 2008), even though the tail extends from a location of ClpS opposite the N-end signal binding site. Therefore, one model of substrate delivery would be for the tail of ClpS to allosterically stimulate substrate release from itself. This is a "catch and release" mechanism because the released substrate is partitioned between $ClpA_6$ binding and rebinding to ClpS. Alternatively, the N-terminal tail of ClpS may activate $ClpA_6$ to pull on the substrate and dislodge it from ClpS. This model involves a direct handoff of substrate from ClpS to $ClpA_6$ and implies that $ClpA_6$ does not initially bind to the N-end residue during delivery.

Both of these delivery models can explain the enhancement of degradation at low substrate concentrations and the suppression of V_{max} by ClpS (chapter 2). In the "catch and release" scheme, ClpS increases the local concentration of substrate near $ClpA_6$ but may cause unproductive steric hindrance of substrate for the binding sites on $ClpA_6$ when substrate is abundant. In other words, $ClpA_6$ is able to capture N-end signals at high substrate concentrations without the aid of ClpS, but the presence of ClpS bound to $ClpA_6$ actually interferes with direct substrate recognition by $ClpA_6$. This is a plausible explanation given the possibility that six ClpS monomers are able to dock onto each ClpA hexamer. To be consistent with previous kinetic data (chapter 2), the direct handoff model assumes that ClpS must first bind substrate and delivery it to $ClpA_6$ even at high substrate concentrations. Slower degradation of N-end substrates at high substrate concentrations may occur because substrate transfer from ClpS to $ClpA_6$ becomes rate-limiting in the degradation mechanism.

Direct interactions between $ClpA_6$ and ClpS may facilitate substrate release from ClpS

A third delivery mechanism was presented at the end of the results in chapter 4 where $ClpA_6$ interacts directly with ClpS to destabilize the hydrogen bonding network used to recognize the α-amino group of the N-end signal. This activity by $ClpA_6$ would release the substrate from ClpS and allow $ClpA_6$ to engage the N-end signal. Disruption of Tyr 37 and Glu 41, two conserved positions in *E. coli* ClpS, slows degradation of an N-end GFP substrate (Erbse et al. 2006). Both residues are near the N-end binding pocket but have no obvious role in N-end signal recognition based on our cocrystal structure. If mutation of Y37A and E41A does not affect binding of ClpS to N-end signals but affects the K_M or V_{max} of N-end substrate degradation by ClpAPS, these residues could be important for interactions with $ClpA_6$ that promote substrate release from ClpS.

References

Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D. Biol Crystallogr* **58**, 1948 (2002).

Bachmair, A., Finley, D., and Varshavsky, A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**(4773): 179-186.

Baker, R.T. and Varshavsky, A. 1991. Inhibition of the N-end rule pathway in living cells. *Proc Natl Acad Sci U S A* **88**(4): 1090-1094.

Baker, T.A. and Sauer, R.T. 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**(12): 647-653.

Balzi, E., Choder, M., Chen, W.N., Varshavsky, A., and Goffeau, A. 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of Saccharomyces cerevisiae. *The Journal of biological chemistry* **265**(13): 7464-7471.

Bartel, B., Wunning, I., and Varshavsky, A. 1990. The recognition component of the N-end rule pathway. *Embo J* **9**(10): 3179-3189.

Benaroudj, N., Tarcsa, E., Cascio, P., and Goldberg, A.L. 2001. The unfolding of substrates and ubiquitin-independent protein degradation by proteasomes. *Biochimie* **83**(3-4): 311-318.

Bolon, D.N., Grant, R.A., Baker, T.A., and Sauer, R.T. 2004. Nucleotide-dependent substrate handoff from the SspB adaptor to the AAA+ ClpXP protease. *Mol Cell* **16**(3): 343-350.

Bota, D.A. and Davies, K.J. 2002. Lon protease preferentially degrades oxidized mitochondrial aconitase by an ATP-stimulated mechanism. *Nat Cell Biol* **4**(9): 674-680.

Bowie, J.U. and Sauer, R.T. 1989. Identification of C-terminal extensions that protect proteins from intracellular proteolysis. *The Journal of biological chemistry* **264**(13): 7596-7602.

Bruick, R.K. and McKnight, S.L. 2001. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**(5545): 1337-1340.

Byrd, C., Turner, G.C., and Varshavsky, A. 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *Embo J* **17**(1): 269-277.

Chen, X., Barton, L.F., Chi, Y., Clurman, B.E., and Roberts, J.M. 2007. Ubiquitin-independent degradation of cell-cycle inhibitors by the REGgamma proteasome. *Mol Cell* **26**(6): 843-852.

Chin, D.T., Goff, S.A., Webster, T., Smith, T., and Goldberg, A.L. 1988. Sequence of the lon gene in Escherichia coli. A heat-shock gene which encodes the ATP-dependent protease La. *The Journal of biological chemistry* **263**(24): 11718-11728.

Craig, K.L. and Tyers, M. 1999. The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog Biophys Mol Biol* **72**(3): 299-328.

Cranz-Mileva, S., Imkamp, F., Kolygo, K., Maglica, Z., Kress, W., and Weber-Ban, E. 2008. The Flexible Attachment of the N-Domains to the ClpA Ring Body Allows their Use On Demand. *J Mol Biol.*

Crews, C.M. 2003. Feeding the machine: mechanisms of proteasome-catalyzed degradation of ubiquitinated proteins. *Curr Opin Chem Biol* **7**(5): 534-539.

Dalboge, H., Bayne, S., and Pedersen, J. 1990. In vivo processing of N-terminal methionine in E. coli. *FEBS Lett* **266**(1-2): 1-3.

Daniel, R.M., Dines, M., and Petach, H.H. 1996. The denaturation and degradation of stable enzymes at high temperatures. *Biochem J* **317 (Pt 1)**: 1-11.

de Groot, R.J., Rumenapf, T., Kuhn, R.J., Strauss, E.G., and Strauss, J.H. 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc Natl Acad Sci U S A* **88**(20): 8967-8971.

Ditzel, M., Wilson, R., Tenev, T., Zachariou, A., Paul, A., Deas, E., and Meier, P. 2003. Degradation of DIAP1 by the N-end rule pathway is essential for regulating apoptosis. *Nat Cell Biol* **5**(5): 467-473.

Dohmen, R.J., Madura, K., Bartel, B., and Varshavsky, A. 1991. The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. *Proc Natl Acad Sci U S A* **88**(16): 7351-7355.

Dougan, D.A., Reid, B.G., Horwich, A.L., and Bukau, B. 2002. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**(3): 673-683.

Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T., and Varshavsky, A. 2002. Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain. *Proc Natl Acad Sci U S A* **99**(22): 14110-14115.

Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D.A., and Bukau, B. 2006. ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**(7077): 753-756.

Farrell, C.M., Baker, T.A., and Sauer, R.T. 2007. Altered specificity of a AAA+ protease. *Mol Cell* **25**(1): 161-166.

Feldman, R.M., Correll, C.C., Kaplan, K.B., and Deshaies, R.J. 1997. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**(2): 221-230.

Flinta, C., Persson, B., Jornvall, H., and von Heijne, G. 1986. Sequence determinants of cytosolic N-terminal protein processing. *Eur J Biochem* **154**(1): 193-196.

Flynn, J.M., Levchenko, I., Seidel, M., Wickner, S.H., Sauer, R.T., and Baker, T.A. 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **98**(19): 10584-10589.

Flynn, J.M., Neher, S.B., Kim, Y.I., Sauer, R.T., and Baker, T.A. 2003. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**(3): 671-683.

Gonda, D.K., Bachmair, A., Wunning, I., Tobias, J.W., Lane, W.S., and Varshavsky, A. 1989. Universality and structure of the N-end rule. *The Journal of biological chemistry* **264**(28): 16700-16712.

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

Gottesman, S., Clark, W.P., de Crecy-Lagard, V., and Maurizi, M.R. 1993. ClpX, an alternative subunit for the ATP-dependent Clp protease of Escherichia coli. Sequence and in vivo activities. *The Journal of biological chemistry* **268**(30): 22618-22626.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**(9): 1338-1347.

Guo, F., Esser, L., Singh, S.K., Maurizi, M.R., and Xia, D. 2002a. Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *The Journal of biological chemistry* **277**(48): 46753-46762.

Guo, F., Maurizi, M.R., Esser, L., and Xia, D. 2002b. Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *The Journal of biological chemistry* **277**(48): 46743-46752.

Hamilton, M.H., Cook, L.A., McRackan, T.R., Schey, K.L., and Hildebrandt, J.D. 2003. Gamma 2 subunit of G protein heterotrimer is an N-end rule ubiquitylation substrate. *Proc Natl Acad Sci U S A* **100**(9): 5081-5086.

Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P. 2007. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell* **26**(1): 131-143.

Herman, C., Thevenet, D., D'Ari, R., and Bouloc, P. 1995. Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. *Proc Natl Acad Sci U S A* **92**(8): 3516-3520.

Hershko, A. 1997. Roles of ubiquitin-mediated proteolysis in cell cycle control. *Curr Opin Cell Biol* **9**(6): 788-799.

Hinnerwisch, J., Fenton, W.A., Furtak, K.J., Farr, G.W., and Horwich, A.L. 2005a. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121**(7): 1029-1041.

Hinnerwisch, J., Reid, B.G., Fenton, W.A., and Horwich, A.L. 2005b. Roles of the N-domains of the ClpA unfoldase in binding substrate proteins and in stable complex formation with the ClpP protease. *The Journal of biological chemistry* **280**(49): 40838-40844.

Hon, W.C., Wilson, M.I., Harlos, K., Claridge, T.D., Schofield, C.J., Pugh, C.W., Maxwell, P.H., Ratcliffe, P.J., Stuart, D.I., and Jones, E.Y. 2002. Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature* **417**(6892): 975-978.

Hoskins, J.R., Singh, S.K., Maurizi, M.R., and Wickner, S. 2000. Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proc Natl Acad Sci U S A* **97**(16): 8892-8897.

Hou, J.Y., Sauer, R.T., and Baker, T.A. 2008. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol* **15**(3): 288-294.

Hoyt, M.A. and Coffino, P. 2004. Ubiquitin-free routes into the proteasome. *Cell Mol Life Sci* **61**(13): 1596-1600.

Hu, R.G., Sheng, J., Qi, X., Xu, Z., Takahashi, T.T., and Varshavsky, A. 2005. The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* **437**(7061): 981-986.

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J., Frishman, D. 2008. Protein abundance profiling of the *Escherchia coli* cytosol. *BMC Genomics* **9**: 102-118.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G., Jr. 2001. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**(5516): 464-468.

Kast, P. 1994. pKSS--a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138**(1-2): 109-114.

Kast, P. and Hennecke, H. 1991. Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *J Mol Biol* **222**(1): 99-124.

Katayama, Y., Gottesman, S., Pumphrey, J., Rudikoff, S., Clark, W.P., and Maurizi, M.R. 1988. The two-component, ATP-dependent Clp protease of Escherichia coli. Purification, cloning, and mutational analysis of the ATP-binding component. *The Journal of biological chemistry* **263**(29): 15226-15236.

Keiler, K.C., Waller, P.R., and Sauer, R.T. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**(5251): 990-993.

Kenniston, J.A., Baker, T.A., Fernandez, J.M., and Sauer, R.T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**(4): 511-520.

Kim, Y.I., Burton, R.E., Burton, B.M., Sauer, R.T., and Baker, T.A. 2000. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**(4): 639-648.

Kwon, Y.T., Kashina, A.S., Davydov, I.V., Hu, R.G., An, J.Y., Seo, J.W., Du, F., and Varshavsky, A. 2002. An essential role of N-terminal arginylation in cardiovascular development. *Science* **297**(5578): 96-99.

Lam, Y.A., Lawson, T.G., Velayutham, M., Zweier, J.L., and Pickart, C.M. 2002. A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. *Nature* **416**(6882): 763-767.

Lam, Y.A., Xu, W., DeMartino, G.N., and Cohen, R.E. 1997. Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature* **385**(6618): 737-740.

Lee, M.J., Tasaki, T., Moroi, K., An, J.Y., Kimura, S., Davydov, I.V., and Kwon, Y.T. 2005. RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. *Proc Natl Acad Sci U S A* **102**(42): 15030-15035.

Levchenko, I., Grant, R.A., Wah, D.A., Sauer, R.T., and Baker, T.A. 2003. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol Cell* **12**(2): 365-372.

Levchenko, I., Seidel, M., Sauer, R.T., and Baker, T.A. 2000. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**(5488): 2354-2356.

Li, S.J. and Hochstrasser, M. 1999. A new protease required for cell-cycle progression in yeast. *Nature* **398**(6724): 246-251.

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

Lloyd, A.G., Ng, Y.S., Muesing, M.A., Simon, V., and Mulder, L.C. 2007. Characterization of HIV-1 integrase N-terminal mutant viruses. *Virology* **360**(1): 129-135.

Lo, J.H., Baker, T.A., and Sauer, R.T. 2001. Characterization of the N-terminal repeat domain of Escherichia coli ClpA-A class I Clp/HSP100 ATPase. *Protein Sci* **10**(3): 551-559.

Lupas, A.N. and Koretke, K.K. 2003. Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. *J Struct Biol* **141**(1): 77-83.

Madura, K. and Varshavsky, A. 1994. Degradation of G alpha by the N-end rule pathway. *Science* **265**(5177): 1454-1458.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., and Butt, T.R. 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **5**(1-2): 75-86.

Martin, A., Baker, T.A., and Sauer, R.T. 2008. Diverse Pore Loops of the AAA+ ClpX Machine Mediate Unassisted and Adaptor-Dependent Recognition of ssrA-Tagged Substrates. *Mol Cell* **29**(4): 441-450.

Maurizi, M.R., Clark, W.P., Kim, S.H., and Gottesman, S. 1990. Clp P represents a unique family of serine proteases. *The Journal of biological chemistry* **265**(21): 12546-12552.

Min, J.H., Yang, H., Ivan, M., Gertler, F., Kaelin, W.G., Jr., and Pavletich, N.P. 2002. Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling. *Science* **296**(5574): 1886-1889.

Mizushima, T., Hirao, T., Yoshida, Y., Lee, S.J., Chiba, T., Iwai, K., Yamaguchi, Y., Kato, K., Tsukihara, T., and Tanaka, K. 2004. Structural basis of sugar-recognizing ubiquitin ligase. *Nat Struct Mol Biol* **11**(4): 365-370.

Muffler, A., Fischer, D., Altuvia, S., Storz, G., and Hengge-Aronis, R. 1996. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in Escherichia coli. *Embo J* **15**(6): 1333-1339.

Mulder, L.C. and Muesing, M.A. 2000. Degradation of HIV-1 integrase by the N-end rule pathway. *The Journal of biological chemistry* **275**(38): 29749-29753.

Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F.B., Mendenhall, M.D., Sicheri, F., Pawson, T., and Tyers, M. 2001. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**(6863): 514-521.

Neher, S.B., Flynn, J.M., Sauer, R.T., and Baker, T.A. 2003. Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes Dev* **17**(9): 1084-1089.

Neidhardt, F.C., and Umbarger, H.E. "Chemical composition of *E. coli* ," *Escherichia coli* and *Salmonella*, p. 13-16, 1996, F.C. Neidhardt, Ed. American Society for Microbiology Press (Washington, DC).

Nijman, S.M., Luna-Vargas, M.P., Velds, A., Brummelkamp, T.R., Dirac, A.M., Sixma, T.K., and Bernards, R. 2005. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**(5): 773-786.

Orlicky, S., Tang, X., Willems, A., Tyers, M., and Sicheri, F. 2003. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112**(2): 243-256.

Otwinowski, Z., and Minor, W. "Processing of X-ray Diffraction Data Collected in Oscillation Mode," Methods in Enzymology, Volume 276: Macromolecular

Crystallography, part A, p. 307-326, 1997, C. W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).

Parag, H.A., Raboy, B., and Kulka, R.G. 1987. Effect of heat shock on protein degradation in mammalian cells: involvement of the ubiquitin system. *Embo J* **6**(1): 55-61.

Parsell, D.A. and Lindquist, S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu Rev Genet* **27**: 437-496.

Parsell, D.A., Silber, K.R., and Sauer, R.T. 1990. Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev* **4**(2): 277-286.

Passmore, L.A. and Barford, D. 2004. Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochem J* **379**(Pt 3): 513-525.

Peters, J.M. 2002. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol Cell* **9**(5): 931-943.

Pickart, C.M. 1997. Targeting of substrates to the 26S proteasome. *Faseb J* **11**(13): 1055-1066.

Pickart, C.M. 2001. Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**: 503-533.

Piszczek, G., Rozycki, J., Singh, S.K., Ginsburg, A., and Maurizi, M.R. 2005. The molecular chaperone, ClpA, has a single high affinity peptide binding site per hexamer. *The Journal of biological chemistry* **280**(13): 12221-12230.

Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E., and Matouschek, A. 2004. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **11**(9): 830-837.

Pratt, L.A. and Silhavy, T.J. 1996. The response regulator SprE controls the stability of RpoS. *Proc Natl Acad Sci U S A* **93**(6): 2488-2492.

Pruteanu, M. and Hengge-Aronis, R. 2002. The cellular level of the recognition factor RssB is rate-limiting for sigmaS proteolysis: implications for RssB regulation and signal transduction in sigmaS turnover in Escherichia coli. *Mol Microbiol* **45**(6): 1701-1713.

Pruteanu, M., Neher, S.B., and Baker, T.A. 2007. Ligand-controlled proteolysis of the Escherichia coli transcriptional regulator ZntR. *J Bacteriol* **189**(8): 3017-3025.

Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. 2001. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410**(6831): 955-959.

Rohrwild, M., Coux, O., Huang, H.C., Moerschell, R.P., Yoo, S.J., Seol, J.H., Chung, C.H., and Goldberg, A.L. 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**(12): 5808-5813.

Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., Neher, S.B., Oakes, E.S., Siddiqui, S.M., Wah, D.A., and Baker, T.A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

Savel'ev, A.S., Novikova, L.A., Kovaleva, I.E., Luzikov, V.N., Neupert, W., and Langer, T. 1998. ATP-dependent proteolysis in mitochondria. m-AAA protease and PIM1

protease exert overlapping substrate specificities and cooperate with the mtHsp70 system. *The Journal of biological chemistry* **273**(32): 20596-20602.

Scheffner, M., Nuber, U., and Huibregtse, J.M. 1995. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**(6509): 81-83.

Schirmer, E.C., Glover, J.R., Singer, M.A., and Lindquist, S. 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci* **21**(8): 289-296.

Seol, J.H., Baek, S.H., Kang, M.S., Ha, D.B., and Chung, C.H. 1995. Distinctive roles of the two ATP-binding sites in ClpA, the ATPase component of protease Ti in Escherichia coli. *The Journal of biological chemistry* **270**(14): 8087-8092.

Siddiqui, S.M., Sauer, R.T., and Baker, T.A. 2004. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev* **18**(4): 369-374.

Singh, S.K., Grimaud, R., Hoskins, J.R., Wickner, S., and Maurizi, M.R. 2000. Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP. *Proc Natl Acad Sci U S A* **97**(16): 8898-8903.

Singh, S.K. and Maurizi, M.R. 1994. Mutational analysis demonstrates different functional roles for the two ATP-binding sites in ClpAP protease from Escherichia coli. *The Journal of biological chemistry* **269**(47): 29537-29545.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. 1997. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**(2): 209-219.

Song, H.K. and Eck, M.J. 2003. Structural basis of degradation signal recognition by SspB, a specificity-enhancing factor for the ClpXP proteolytic machine. *Mol Cell* **12**(1): 75-86.

Sousa, M.C., Trame, C.B., Tsuruta, H., Wilbanks, S.M., Reddy, V.S., and McKay, D.B. 2000. Crystal and solution structures of an HslUV protease-chaperone complex. *Cell* **103**(4): 633-643.

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

Takeuchi, J., Chen, H., Hoyt, M.A., and Coffino, P. 2008. Structural elements of the ubiquitin-independent proteasome degron of ornithine decarboxylase. *Biochem J* **410**(2): 401-407.

Tasaki, T., Mulder, L.C., Iwamatsu, A., Lee, M.J., Davydov, I.V., Varshavsky, A., Muesing, M., and Kwon, Y.T. 2005. A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**(16): 7120-7136.

Tasaki, T., Sohr, R., Xia, Z., Hellweg, R., Hortnagl, H., Varshavsky, A., and Kwon, Y.T. 2007. Biochemical and genetic studies of UBR3, a ubiquitin ligase with a function in olfactory and other sensory systems. *The Journal of biological chemistry* **282**(25): 18510-18520.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. 2000. Recognition of the polyubiquitin proteolytic signal. *Embo J* **19**(1): 94-102.

Tobias, J.W., Shrader, T.E., Rocap, G., and Varshavsky, A. 1991. The N-end rule in bacteria. *Science* **254**(5036): 1374-1377.

Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A.J., Oppenheim, A.B., Yura, T., Yamanaka, K., Niki, H., 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**(11): 2551-2560.

Tu, G.F., Reid, G.E., Zhang, J.G., Moritz, R.L., and Simpson, R.J. 1995. C-terminal extension of truncated recombinant proteins in Escherichia coli with a 10Sa RNA decapeptide. *The Journal of biological chemistry* **270**(16): 9322-9326.

Turgay, K., Hahn, J., Burghoorn, J., and Dubnau, D. 1998. Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. *Embo J* **17**(22): 6730-6738.

Varshavsky, A. 1992. The N-end rule. *Cell* **69**(5): 725-735.

Varshavsky, A. 1996. The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci U S A* **93**(22): 12142-12149.

Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., 3rd, Koonin, E.V., and Deshaies, R.J. 2002. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**(5593): 611-615.

Verma, R., Oania, R., Graumann, J., and Deshaies, R.J. 2004. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell* **118**(1): 99-110.

Visick, J.E. and Clarke, S. 1995. Repair, refold, recycle: how bacteria can deal with spontaneous and environmental damage to proteins. *Mol Microbiol* **16**(5): 835-845.

Wah, D.A., Levchenko, I., Rieckhof, G.E., Bolon, D.N., Baker, T.A., and Sauer, R.T. 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**(2): 355-363.

Walker, G.C., Marsh, L., and Dodson, L. 1985. Cellular responses to DNA damage. *Environ Health Perspect* **62**: 115-117.

Wang, J., Hartling, J.A., and Flanagan, J.M. 1997. The structure of ClpP at 2.3 A resolution suggests a model for ATP- dependent proteolysis. *Cell* **91**(4): 447-456.

Wang, K.H., Sauer, R.T., and Baker, T.A. 2007. ClpS modulates but is not essential for bacterial N-end rule degradation. *Genes Dev* **21**(4): 403-408.

Wang, K.H., Oakes, E.S.C., Sauer, R.T., and Baker, T.A. *J Biol Chem, accepted.*

Weber-Ban, E.U., Reid, B.G., Miranker, A.D., and Horwich, A.L. 1999. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**(6748): 90-93.

Willam, C., Masson, N., Tian, Y.M., Mahmood, S.A., Wilson, M.I., Bicknell, R., Eckardt, K.U., Maxwell, P.H., Ratcliffe, P.J., and Pugh, C.W. 2002. Peptide blockade of HIFalpha degradation modulates cellular metabolism and angiogenesis. *Proc Natl Acad Sci U S A* **99**(16): 10423-10428.

Xia, D., Esser, L., Singh, S.K., Guo, F., and Maurizi, M.R. 2004. Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *J Struct Biol* **146**(1-2): 166-179.

Yoshida, Y., Chiba, T., Tokunaga, F., Kawasaki, H., Iwai, K., Suzuki, T., Ito, Y., Matsuoka, K., Yoshida, M., Tanaka, K., and Tai, T. 2002. E3 ubiquitin ligase that recognizes sugar chains. *Nature* **418**(6896): 438-442.

Yoshida, Y., Tokunaga, F., Chiba, T., Iwai, K., Tanaka, K., and Tai, T. 2003. Fbs2 is a new member of the E3 ubiquitin ligase family that recognizes sugar chains. *The Journal of biological chemistry* **278**(44): 43877-43884.

Zeth, K., Ravelli, R.B., Paal, K., Cusack, S., Bukau, B., and Dougan, D.A. 2002. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat Struct Biol* **9**(12): 906-911.