Translocation and proteolysis by the energy-dependent protease ClpAP: Coordination of conformational changes and active site chemistry

by

Laura Danielle Jennings

B.S., Chemistry and Mathematics (2004)
University of Missouri-Columbia

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in Biological Chemistry

at the

Massachusetts Institute of Technology

September 2008

©2008 Massachusetts Institute of Technology
All rights reserved

Signature of Author  

Department of Chemistry
July 29, 2008

Certified by  

Stuart Licht
Assistant Professor of Chemistry
Thesis Supervisor

Accepted by  

Robert W. Field
Chairman, Departmental Committee on Graduate Students
This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor Barbara Imperiali ____________________________ Committee Chair

Professor Stuart Licht ________________________________ Research Supervisor

Professor JoAnne Stubbe ______________________________ Committee Member
Translocation and proteolysis by the energy-dependent protease ClpAP: Coordination of conformational changes and active site chemistry

by

Laura Danielle Jennings

Submitted to the Department of Chemistry on July 29, 2008 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biological Chemistry

ABSTRACT

Energy-dependent proteases, such as the *E. coli* protease ClpAP, degrade misfolded and short-lived regulatory proteins inside the cell. In this class of proteins, an ATPase component (e.g., ClpA) binds, unfolds, and translocates protein substrates into a protease component (e.g., ClpP) where degradation occurs. This thesis addresses the communication between ClpA and ClpP required to efficiently translocate and degrade protein substrates. Synchrotron hydroxyl radical footprinting is used to show that the ClpA D2 loop, located in the pore of ClpA, is in an “up” conformation when ATP is bound and assumes a “down” conformation when ADP is bound. These results provide the first direct experimental support for a nucleotide-dependent D2 loop conformational change previously proposed to mediate substrate translocation. Synchrotron footprinting also shows that the ClpP N-terminal loops undergo a conformational change, upon the binding of ClpA, from a closed, pore-blocking position, to an open, pore-free position. Complementary kinetic studies show that deletion of the ClpP N-terminus (ClpPΔN) accelerates the degradation rate of large peptide substrates 5-15 fold, indicating that opening of the pore is functionally important. Furthermore, unlike ClpAP and wild-type ClpP, ClpPΔN shows a distinct slow phase of product formation that is eliminated with the addition of hydroxylamine, suggesting that deletion of the N-terminal loops stabilizes the acyl-enzyme intermediate. Additionally, size-exclusion chromatography and kinetic studies are used to show that ClpP alone can processively degrade a full-length protein substrate in the absence of ClpA, albeit at a 2000-fold slower rate, and furthermore that the size distribution of ClpP-generated peptide products is strikingly similar to the size distribution of ClpAP-generated peptide products. Both distributions contain peaks at integral multiples of 7-8 amino acids, consistent with a mechanism in which ClpP controls product sizes by alternating between translocation in steps of 7-8 amino acids and proteolysis. Collectively, the results presented here indicate that 1) conformational changes in both ClpA and ClpP are necessary for efficient translocation and proteolysis, 2) active site reactivity is linked to conformational changes in the pore region of ClpP, and 3) product sizes are largely controlled by ClpP.

Thesis Supervisor: Stuart Licht

Title: Assistant Professor of Chemistry
ACKNOWLEDGMENTS

Many people deserve recognition for their contributions to my graduate education. First, I would like to thank my research advisor, Stuart. His creativity, scientific rigor, and enthusiasm for science have been an inspiration to me. My research would not have been successful without Stuart’s willingness to sit down with me on a regular basis and discuss the details of my work. Having Stuart as an advisor has made graduate school a highly positive experience for me, and I cannot thank him enough for his support and encouragement.

I would also like to thank the members of my thesis committee, Prof. Barbara Imperiali, Prof. JoAnne Stubbe, and Prof. Dan Kemp, for their continuous support and guidance.

In addition, I am grateful to have been involved in two highly efficient and successful collaborations. The HPLC deconvolution work described in this thesis was performed in collaboration with Prof. Muriel Médard (MIT, Department of Electrical Engineering and Computer Science) and Dr. Desmond Lun (formerly of the MIT Department of Electrical Engineering and Computer Science, now of the Broad Institute). The synchrotron footprinting work described here was performed in collaboration with Prof. Mark Chance and Dr. Jen Bohon of Case Western Reserve University. Additionally, I’d like to thank Prof. Barbara Imperiali for allowing the use of her HPLC with fluorescence detection.

I’d also like to thank my friends on the 5th floor (known to some as the Fifth Floor Fashizzles, the Penterrors, or the Merohedral Twins) for their scientific guidance and, most of all, their friendship. In particular I’d like to thank my labmates—Mathew for his friendship and knowledge of all things biochemistry, Mary for being my partner in crime on ClpAP, Kee-Hyun for sharing several ClpAP protocols with me, Melva for her friendship and always making me laugh at inappropriate times, and Wan-Chen for sharing her organic chemistry expertise and for our daily greetings. Also, if your labmates are your siblings in graduate school, the Drennan and Klibanov lab members have become my step-siblings. Specifically, I’d like to thank Christine, LauraP, Yan, Leah, Ainsley, Jen, and Alisha for their friendship. I’d like to also extend a special thanks to Eugene. His love and support have been instrumental in my successful completion of graduate school.

Finally, I’d like to thank my parents, Dan and Dianne Jennings. They instilled in me a belief, from a young age, that I could accomplish my dreams, and it has only been through their love, support, and encouragement that this has been true. Mom and Dad, this is for you.
TABLE OF CONTENTS

Chapter I: Introduction to energy-dependent proteases 7
A. Structural features of energy-dependent proteases 7
B. Peptide degradation signals recognized by ClpA 14
C. Unfolding and translocation by ClpA and ClpX are ATP-hydrolysis dependent 18
D. A model for ATP-dependent unfolding and translocation 19
E. Substrate binding by ClpA 22
F. Conformational changes of ClpA involved in substrate translocation 23
G. Structural elements important in communication between ClpA/X and ClpP 26
H. Processivity of ClpAP 30
I. Control of product sizes by ClpAP 31
J. Contents of the thesis 33
K. References 36

Chapter II: The ClpP N-terminus coordinates substrate access with protease active site reactivity 45
A. Summary 45
B. Introduction 46
C. Results 49
D. Discussion 67
E. Materials and Methods 76
F. Acknowledgments 79
G. References 79

Chapter III: ClpP hydrolyzes a protein substrate processively and independently of the ClpA ATPase: Mechanistic studies of ATP-independent processive proteolysis 85
A. Summary 85
B. Introduction 86
C. Results 90
D. Discussion 107
E. Materials and Methods 116
F. Acknowledgments 122
G. References 123
Chapter IV. Synchrotron protein footprinting supports substrate translocation by ClpA via ATP-induced movements of the D2 loop

A. Summary 127
B. Introduction 128
C. Results 134
D. Discussion 145
E. Materials and Methods 151
F. Acknowledgments 154
G. References 155

Appendix A: ClpA purification protocol 159

Appendix B: ClpP-His6 purification protocol 169

Appendix C: GFP-ssrA purification protocol 177

Curriculum Vitae 181
Chapter I: Introduction to energy-dependent proteases

Energy-dependent proteases play a vital role in maintaining cellular homeostasis by degrading misfolded and short-lived regulatory proteins inside the cell [1, 2]. This class of enzymes is conserved across all kingdoms of life, and although the complexity of these proteases varies across species, the overall architecture is conserved. In order to prevent unintended degradation, the protease active sites of energy-dependent proteases are sequestered in barrel-shaped compartments only accessible via narrow entry channels [3, 4]. Protein substrates reach these protease active sites through the action of partner ATPases [5-8] which stack coaxially with the compartmental proteases. These ATPase components bind, unfold, and translocate protein substrates into the proteolytic chamber where degradation occurs [9-11] (Figure I.1).

In eukaryotes, energy-dependent degradation is primarily carried out by the proteasome [3, 4, 12, 13]. However, while proteasomes are present in archaea, although nonessential, they are rare in bacteria. Instead, bacteria contain five other energy-dependent proteases, ClpAP, ClpXP, Lon, HslUV, and FtsH [1, 14]. This thesis will focus mainly on the enzymology of the *E. coli* energy-dependent protease, ClpAP.

I.A. Structural features of energy-dependent proteases

All energy-dependent proteases consist of ring-shaped ATPase components that stack on top of ring-shaped protease components. However, there are some basic structural differences between members of the family. The ATPase component and the protease component can reside on the same polypeptide chain (e.g., Lon and FtsH) or on different chains (e.g., ClpAP, ClpXP, HslUV, the proteasome) [4, 11]. In addition, the
General mechanism of energy-dependent proteases.
ATPase and protease components can be homomeric or heteromeric rings. All prokaryotic energy-dependent proteases (including the archaeal proteasome) are composed of homomeric ring-shaped components, while the eukaryotic proteasome is composed of heteromeric rings and contains >30 different proteins in total [4]. Due to the structural complexity of the eukaryotic proteasome, prokaryotic systems (including ClpAP) are often used as model enzymes to study this class of proteins [10, 11].

I.A.1. Structure of ClpA

The ATPase components of energy-dependent proteases are composed of AAA+ (ATPase associated with various cellular activities) modules [5]. AAA+ ATPases make up a large class of enzymes responsible for a wide variety of cellular activities, including protein repair and degradation, membrane fusion, cell cycle regulation, and DNA replication. Members of this superfamily convert the energy of ATP binding and hydrolysis into mechanical forces used to direct the remodeling of bound substrates [6-8]. AAA+ ATPases generally assemble into oligomeric, ring-shaped structures. In most energy-dependent proteases, the AAA+ ATPase subunits form hexameric rings [4].

The energy-dependent protease ClpAP is composed of the ATPase component ClpA and the protease component ClpP. ClpP can also bind the ATPase component ClpX, forming ClpXP. ClpA and ClpX are structurally similar. Both enzymes form ring-shaped homohexamers in the presence of either ATP or nonhydrolyzable analogs (such as ATPγS or AMPPNP). However, ClpA contains two AAA+ ATPase modules per monomer unit, while ClpX contains only one. This leads to a hexameric ClpA that is composed of essentially two tiers (one per ATPase module), while hexameric ClpX has
only one tier (Figure 1.2) [15]. The functional significance of this difference is not yet clear. For both ClpA [16] and ClpX [17], crystal structures are only available for the monomeric subunits; no hexameric crystal structures have yet been published. However, electron micrographs are available showing both ClpA and ClpX hexamers (Figure 1.2) [15, 18-20]. In addition, hexameric models for both ClpA [16, 21] and ClpX [17] have been published; one such ClpA model is shown in Figure I.3A and I.3B. Both the ClpA and ClpX hexamers have narrow (~10-15 Å) pores through which protein substrates are translocated. ClpA is composed of three distinct domains: a flexible N-terminal domain important for binding unfolded protein substrates [22-24], the D1 ATPase domain implicated in substrate binding and hexamerization [21, 25], and the D2 ATPase domain thought to be important in binding and translocating substrate proteins [21].

I.A.2. Structure of ClpP

The proteolytic active sites of energy-dependent proteases are generally contained in hollow, barrel-shaped compartments. These compartments stack co-axially with ring-shaped AAA+ partners to create a continuous pore through which substrate proteins can be passed sequentially from the ATPase component into the protease component. In the case of ClpP, this barrel-shaped structure is composed of monomeric units assembled into ring-shaped homoheptamers, which stack face-to-face to form a hollow barrel-shaped structure with ~10 Å pores at either end. ClpP’s fourteen serine protease active sites (one per monomeric unit) are contained within this barrel. A number of crystal structures of ClpP are available [26-33]. One such crystal structure is shown in Figure I.3C and I.3D. The ClpP active sites are arranged in two rings, one per heptameric unit. The active sites
**ClpA hexameric model and ClpP crystal structure.** A) Top view of the ClpA hexameric model. The model was constructed using the ClpA monomeric crystal structure and the hexameric crystal structures of NSF-D2 and HslU [16]. The N-terminal domain is in magenta, the D1 ATPase domain is in green, and the D2 ATPase domain is in blue. B) Side view of the ClpA hexameric model. C) Top view of the ClpP crystal structure, PDB accession code 1YG6 [29]. The N-terminal residues are in blue, and the serine active sites are in green. D) Side view of the ClpP crystal structure. “Top” N-terminal residues were crystallized in the “up” conformation. “Bottom” N-terminal residues were crystallized in the “down” conformation [29].
in each ring are ~25 Å apart and are connected via a hydrophobic substrate binding groove [26, 34].

Substrates access the ClpP proteolytic chamber through the axial pores, which are primarily composed of the N-terminal region of ClpP (Figure I.3C and I.3D). The ClpP axial pores align with the axial pores of its partner ATPase components (ClpA or ClpX) to form ClpAP or ClpXP complexes. Since ClpA and ClpX are both hexameric, while ClpP is heptameric, the formation of ClpAP and ClpXP complexes involves a symmetry mismatch between ClpA/X and ClpP [19]. However, how this symmetry mismatch is accommodated is not yet known. While crystal structures of either the ClpAP or ClpXP complex are not available, electron micrograph images of these complexes have been published [15, 19, 20] (Figure I.2). Because ClpP is a double ring, it can bind ClpA or ClpX from both the “top” and the “bottom” (Figure I.2), although binding to both faces is not a requirement for activity [35]. The electron micrographs also show that catalytically active mixed complexes can be formed in which ClpP simultaneously binds ClpA on one face and ClpX on the opposite face [15, 20]. It is likely that these mixed complexes also form in vivo [15], but it is not known if ClpA and ClpX can simultaneously translocate substrates into ClpP.

In addition to axial pores, ClpP is also thought to form transient equatorial pores located at the interface between the two ClpP heptameric rings [28, 36] (Figure I.3D). These equatorial pores have been implicated in peptide product release [36], and in vitro they may also allow the entry of small peptide substrates into the chamber (see Chapter 2 of this thesis).
I.B. Peptide degradation signals recognized by ClpA

Proteolysis must be tightly controlled to prevent unintended degradation. To this end, substrate proteins are recognized and distinguished from non-substrate proteins by the ATPase component of energy-dependent proteases. In eukaryotic systems, ubiquitination is the primary mode of substrate recognition [37]. However, bacteria lack a ubiquitin system, and instead protein substrates are recognized by short, unstructured tag sequences usually found at the N- or C-terminus of substrates [10]. In contrast to ClpXP, for which >50 in vivo substrates have been identified which can be grouped into 5 classes based on their degradation tags [38], relatively few ClpAP in vivo substrates are known. One example of a naturally occurring ClpAP substrate is the DNA replication initiator protein RepA, which is known to be recognized by ClpA via a 15-residue sequence on its N-terminus [39]. Recent work has shown that this N-terminal tag is recognized even if it is moved to the C-terminus or placed at an internal site [39, 40], although placement at an internal site also requires an unstructured terminus for efficient degradation [41]. In vitro experiments have shown that unfolded proteins are also good substrates for ClpAP [42]. It is thought that unfolding, which in vivo may be caused by environmental stresses such as heat shock or oxidation, exposes or unstructures tag sequences that are otherwise buried and/or structured [10]. In addition to the RepA tag and tags exposed by protein unfolding, two other classes of ClpA-recognized tags are well characterized—the ssrA tag and N-end rule substrates. These are the most extensively studied sequence tags and will be discussed individually.

I.B.1. SsrA-tagging
In the mid 1990’s, a unique and powerful bacterial degradation tagging system, the ssrA tagging system, was discovered and characterized. In contrast to the naturally encoded tags found in RepA and unfolded proteins, the ssrA tag is added co-translationally to the C-terminus of proteins that become stalled on the ribosome. The discovery of the ssrA tag was not only important in elucidating the \textit{in vivo} biochemistry of protein degradation, but it also allowed for powerful \textit{in vitro} studies. A brief description of the discovery of the ssrA tag along with the mechanism of tagging is outlined below.

In their 1995 paper [43], the Simpson lab reported that overexpression of murine interleukin-6 (mIL-6) in \textit{E. coli} resulted in a small population of C-terminally truncated proteins to which an 11-amino acid tag, AANDENYALAA, had been added to the C-terminus. They further found that addition of this tag was dependent upon the \textit{ssrA} (small stable RNA) gene [44], which encodes the last 10 amino acids of the tag. However, they were unable to detect any mIL-6-tag mRNA chimeric transcripts, suggesting that the tag was not added during RNA synthesis via a co-transcriptional process. Furthermore, they found no evidence for \textit{ssrA}-encoded peptide products or precursor peptides, arguing against a post-translational modification. They concluded that the tag must be added by a previously undescribed biochemical process [43].

In 1996, the Sauer group identified this unique biochemical process as co-translational tagging. They showed that mRNAs lacking a stop codon (which could occur \textit{in vivo} by premature termination of transcription or cleavage by nucleases) become stalled on the ribosome and are released by co-translation of the ssrA RNA, resulting in the addition of the tag previously identified by the Simpson lab [45]. According to the
working model for ssrA tagging, ssrA RNA can fold into a tRNA-like structure and become charged with an alanine. This alanine-charged ssrA RNA then recognizes a ribosome that is stalled at the end of an mRNA lacking a stop codon. The alanine from the ssrA RNA is added to the C-terminal end of the nascent polypeptide chain forming a peptidyl-ssrA RNA bond. The ribosome then releases the damaged mRNA and begins translation on the ssrA RNA. Translation of the ssrA RNA results in the addition of ANDENYALAA, and translation is terminated at a stop codon encoded in the ssrA RNA. This process results in the addition of the “ssrA tag”, AANDENYALAA, to the C-terminal end of proteins that become stalled on the ribosome.

Further studies by the Sauer group showed that these ssrA-tagged proteins are degraded in vivo by ClpAP and ClpXP [46]. Work by Herman et al. showed that FtsH also plays a role in degrading ssrA-tagged proteins [47]. Furthermore, a number of in vitro studies have shown that the addition of the ssrA tag to the C-terminus of almost any protein, including GFP [48] [49], makes that protein a good substrate for ClpXP or ClpAP [46]. This discovery allowed for biochemical experiments elucidating the mechanism of unfolding and translocation by these proteases, which are described in sections I.C and I.D.

I.B.2. N-end rule substrates and ClpS

Another type of degradation tag recognized by ClpAP are those presented on N-end rule substrates. The N-end rule, present in both eukaryotes and prokaryotes, relates the in vivo half-life of a protein to the identity of its N-terminal residue [50, 51]. In eukaryotes, proteins with destabilizing N-terminal residues are ubiquitinated and
degraded by the proteasome [50, 51]. In bacteria, N-end rule substrates are degraded by ClpAP [52].

The primary destabilizing residues in bacteria are F, L, W, and Y. Proteins with these residues at their N-terminus are degraded with an in vivo half-life of approximately 2 minutes [52]. The secondary destabilizing residues in E. coli are R and K. Proteins with these residues at their N-terminus are recognized by a F/L transferase which adds F or L to the N-terminus. Following this modification, these proteins are also rapidly degraded [52].

The degradation of N-end rule substrates in bacteria is enhanced by the ClpA adaptor protein ClpS [53]. Although a number of ClpX adaptor proteins have been characterized [11], ClpS is the only known adaptor for ClpA. ClpS is monomeric, and one ClpS molecule can bind one ClpA monomer, therefore up to six ClpS molecules can bind one ClpA hexamer [54, 55]. ClpS binds to the N-terminal domain of ClpA [54, 55] and has been shown to enhance the degradation of aggregates but slow the degradation of ssrA-tagged substrates [56, 57]. Erbse et al. showed that ClpS binds directly to the N-terminal destabilizing residues of N-end rule substrates and hypothesized that ClpS plays an essential role in vivo by delivering these substrates to ClpA for degradation. In support of this idea, Erbse et al. showed that a subset of N-end rule substrates are stable in ΔclpS cells [53]. In contrast, in 2007, Wang et al. reported that a different subset of N-end rule substrates are degraded in vivo even in ΔclpS cells [58]. Furthermore, this group showed that while ClpS increases the degradation rate of N-end rule substrates at low (<20 μM) concentrations, at high (>75 μM) concentrations of substrate, ClpS has an inhibitory effect [58]. These results are consistent with the idea that ClpS can deliver N-
end rule substrates to ClpA and therefore enhance their degradation at low concentrations; however, it appears that ClpA alone can also recognize these substrates and degrade them efficiently at high concentrations. Therefore, ClpS is not essential for the degradation of N-end rule substrates.

I.C. Unfolding and translocation by ClpA and ClpX are ATP-hydrolysis dependent.

The discovery and characterization of a number of degradation tags recognized by ClpA and ClpX allowed for key in vitro biochemical studies elucidating various mechanistic aspects of these enzymes. Of particular importance is the role of ATP hydrolysis in ClpA and ClpX-catalyzed unfolding and translocation.

In 1999 Weber-Ban et al. showed that ClpA unfolds GFP-ssrA in a reaction dependent upon ATP hydrolysis [48]. While the non-hydrolyzable analog ATPγS allows binding to GFP-ssrA, it does not support unfolding. Importantly, this group showed that ClpA can unfold GFP-ssrA alone, in the absence of ClpP, consistent with previous reports indicating that ClpA by itself acts as a molecular chaperone [59]. These studies showed that ClpA is an ATP-dependent unfoldase, and furthermore that the addition of a recognition tag, such as ssrA, facilitates the unfolding of very stable proteins. A subsequent study by Hoskins et al. in 2000 [42] showed that ClpA can also unfold GFP in an ATP-dependent fashion if the RepA degradation signal is attached to the N-terminus of GFP. As was shown for GFP-ssrA, RepA-GFP can bind ClpA in the presence of nonhydrolyzable ATP analogs but requires ATP hydrolysis for unfolding. Hoskins et al. further showed that ClpAP can degrade this fusion protein in vivo. In addition, this study showed that translocation by ClpA requires ATP hydrolysis. Hoskins
et al. found that ClpA could recognize acid-denatured GFP in the absence of a degradation tag; however, the degradation of this protein required ATP hydrolysis despite the fact that the unfolding step was bypassed. Subsequent electron microscopy studies by Ishikawa et al. showed that ClpA binds protein substrates in the presence of ATPγS, and upon addition of ATP, translocates these substrates directly into the ClpP degradation chamber without large structural rearrangements [60]. Collectively, these studies showed that ClpA catalyzes both the unfolding and translocation of protein substrates, that both steps require ATP hydrolysis, and that these steps occur without major structural rearrangements of the ClpAP complex.

Similar studies were also performed using ClpXP. A 2000 study by Kim et al. showed that ClpX also unfolds GFP-ssrA in an ATP-dependent manner [61]. A study published later that year by Singh et al. confirmed this finding and found in addition that unfolded GFP-ssrA also requires ATP hydrolysis for degradation by ClpXP [49]. Electron microscopy studies similar to those performed for ClpAP showed that ClpXP binds protein substrates at distal sites and, in the presence of ATP, translocates these substrates directly into the ClpP chamber [62]. These findings show that both ClpA and ClpX require ATP hydrolysis for the unfolding and translocation of protein substrates, and further suggest that these two enzymes work via a similar mechanism.

I.D. A model for ATP-dependent unfolding and translocation.

Once it was established that both unfolding and translocation by ClpA and ClpX are ATP hydrolysis-dependent, the Baker and Sauer groups set about defining a model for these two processes. Collectively, their studies have led to a proposed mechanism in
which ClpA or ClpX converts the energy of ATP-hydrolysis into a "pulling" force that drives both unfolding and translocation. Studies by Kenniston et al. [63-65] have shown that unfolding and translocation happen in two steps (Figure I.4). The first is an unfolding/engagement step. In this step, the enzyme "pulls" on the substrate from the degradation tag in an attempt to unfold or partially unfold the substrate. If the enzyme is unsuccessful in unfolding, the substrate can dissociate in which case it must bind again to continue the process. If, however, the enzyme is successful in unfolding or partially unfolding the substrate, the enzyme-substrate complex becomes stabilized, and the substrate is engaged, ready for translocation. After engagement, the enzyme then uses the same mechanical force to translocate the substrate into the ClpP chamber for subsequent degradation [10, 11].

The number of ATPs hydrolyzed during unfolding and translocation has also been quantified, primarily for ClpX. It has been shown by both the Matouschek group and the Baker and Sauer groups that the number of ATP's hydrolyzed during unfolding is highly dependent upon the stability of the protein structure directly adjacent to the degradation tag [63, 64, 66]. If the protein substrate is stable in this region, the enzyme will require more rounds of ATP hydrolysis to unfold and engage the substrate. For hyperstable substrates, the enzyme can use up to 500 molecules of ATP during this step [63]. However, if the substrate has an unstable fold directly adjacent to the degradation tag, many fewer ATP molecules are required for unfolding/engagement. In fact, fewer than 20 molecules of ATP have been reported to be hydrolyzed during this step for highly unstable protein substrates [63]. In contrast, the translocation step appears to be largely independent of the stability of the substrate. It has been reported that ClpX uses 0.2-1
Model for unfolding and translocation by ClpA or ClpX. Unfolding and translocation both occur via a mechanical “pulling” mechanism driven by ATP hydrolysis. The first step is unfolding and engagement of the substrate. ClpA or ClpX exerts a mechanical force on the substrate from the degradation-tagged end. The enzyme is frequently unsuccessful in unfolding the substrate, in which case the substrate is released. Occasionally, the enzyme successfully unfolds the substrate, leading to stabilization of the enzyme-substrate complex (engagement). Following engagement, the substrate is translocated (step 2) into the degradation chamber of ClpP via the same mechanical pulling mechanism. The number of cycles required before engagement depends on the stability of the substrate structure directly adjacent to the degradation tag. Translocation efficiency does not depend on the stability of the substrate protein. Figure adapted from Ref. 10.
molecules of ATP per residue of the substrate during translocation, regardless of the stability of the protein [63, 67].

It is somewhat surprising that ClpA and ClpX use so much energy to fuel protein degradation. In the case of stable protein substrates, these enzymes use as much or more ATP to degrade the substrate than was used to make it [63]. However, two general hypotheses have been put forward to explain these results. The first hypothesis is that protein degradation is vital to cell homeostasis, and therefore, it is not unreasonable that a large amount of energy is devoted to discarding these potentially hazardous proteins. The second argument is that perhaps the “pulling” mechanism (Figure 1.4) is the most efficient mechanism for dealing with a wide variety of protein substrates. ClpA and ClpX must degrade protein substrates with an assortment of structures and stabilities, and therefore these enzymes cannot optimize activity for any single substrate. If ClpA and ClpX were specific unfoldases, a more efficient mechanism may have evolved [10, 11]. It is likely that both explanations will contribute to a complete understanding of the role of ATP hydrolysis in mechanical unfolding and translocation by ClpA and ClpX.

I.E. Substrate binding by ClpA

Studies have shown that only one substrate binding site exists per ClpA hexamer [68], consistent with structural data showing a narrow pore (∼10-15 Å) large enough to accommodate an unfolded polypeptide or perhaps an alpha helix [16, 18, 19]. These substrate proteins are recognized via a number of different sites within ClpA. As discussed in section I.A.1., ClpA consists of three domains—the N-terminal domain, the D1 ATPase domain, and the D2 ATPase domain (Figure I.3). All three domains appear
to play a role in substrate binding. The N-domain, the domain most distal from the ClpP-binding face of ClpA, is necessary for binding RepA [21, 23] and enhances the binding of unfolded proteins [22]. Electron microscopy studies suggest that this domain is flexible; therefore, the N-terminal domain may be important in delivering these substrates to the pore of ClpA for further processing [18]. However, not all substrates require the N-terminal domain for binding. Although some ssrA-tagged substrate proteins are thought to bind the N-domain [21, 23], the ssrA tag itself does not bind the ClpA N-domain. Instead, crosslinking studies indicate that the ssrA tag binds to a pore-facing loop in the D2 domain, deep within the ClpA pore [21] (Figure I.5A). In accordance with these findings, deletion of the N-terminal domain does not affect binding of ssrA-tagged substrates [22, 23, 68]; instead, mutations in pore-facing loop regions of the D1 and D2 domains abolish binding to ssrA-tagged substrates [21]. Interestingly, these same mutations allow binding of RepA-tagged substrates, but do not allow further processing (i.e., unfolding, translocation, or degradation), suggesting that loop regions in the D1 and D2 domains are necessary for binding RepA-tagged substrates during subsequent degradation steps [21]. Furthermore, mutations in D1 and D2 pore-facing loops also abolish binding of unfolded substrates [21], suggesting that efficient binding to these substrates requires elements within the N-domain, the D1 domain, and the D2 domain.

I.F. Conformational changes of ClpA involved in substrate translocation

The fact that mutations in loops within the central channel of ClpA abolish binding and/or degradation of substrates containing multiple types of degradation tags [21] suggests that these substrates share a common translocation path through the ClpA
The ClpA D2 loop is proposed to mediate translocation through the pore. A) Hexameric crystal structure of p97 (PDB code 1OZ4), which has significant structural homology to ClpA. The p97 loop analogous to the ClpA D2 loop (ClpA residues 526-538) is highlighted in green. Crosslinking studies show that the ClpA D2 loop binds the ssrA tag region of ssrA-tagged substrates. Also, mutations in this loop abolish binding and/or degradation of RepA-tagged, ssrA-tagged, and unfolded protein substrates [21]. B) Translocation mechanism proposed by Hinnerwisch et al. [21] ATP binding causes the D2 loop (green) to assume an “up” conformation where it can bind the terminus of a substrate protein (shown here binding the ssrA tag (blue) of GFP-ssrA). ATP hydrolysis causes the loop to move down towards the ClpP binding face of ClpA, dragging the substrate through the central channel of ClpA. This causes simultaneous unfolding and translocation of the substrate. Figure adapted from Ref. 21.
pore. While a number of mutations within the D2 pore-facing loop (residues 526-538, Figure I.5A) abolish binding to ssrA-tagged and unfolded proteins, Hinnerwisch et al. identified one mutation immediately C-terminal to this loop, Y540A, that allows binding of ssrA-tagged, RepA-tagged, and unfolded substrates, but blocks degradation of these substrates [21]. In the ClpA hexameric model [16], Y540 sits at the “hinge” point where the D2 loop inserts into the wall of the ClpA channel. Both the location of this residue and the observation that mutations at this position allow binding but block degradation suggest that the Y540 residue plays a role in controlling the movement of the D2 loop. Furthermore, Hinnerwisch et al. hypothesized that this putative loop movement could be coupled to nucleotide hydrolysis. Y540 sits within a conserved GYVG motif. This motif is also part of a channel-facing loop in the HslU ATPase, and structural studies of this system have shown that the conformation of this loop varies with nucleotide state [69, 70]. A similar mechanism could apply for ClpA. Based on these data and observations, Hinnerwisch et al. put forth a model for translocation through the ClpA pore. They hypothesized that ATP binding places the D2 loop in an “up” conformation in which it can bind the terminus of a substrate protein. ATP hydrolysis then causes the loop to move “down” towards the ClpP binding face of ClpA. This movement of the loop drags the substrate through the ClpA channel causing simultaneous unfolding and translocation of the substrate [21] (Figure I.5B). Chapter 4 of this thesis will directly assess this proposed ATP-dependent loop movement in ClpA using synchrotron hydroxyl radical footprinting.

The mechanism proposed above effectively converts ATP hydrolysis into a mechanical pulling force that allows unfolding and translocation of the substrate protein.
This idea is consistent with the “pulling” mechanism of unfolding and translocation put forth by the Baker and Sauer labs (see section I.D and Figure I.4). In addition, similar ATP-dependent pore loop movements have been hypothesized to drive substrate unfolding and translocation in ClpX. Crosslinking studies show that ClpX also binds ssrA via pore loops, and furthermore that the crosslinking efficiency is dependent upon the nucleotide state of the enzyme [71]. These data show that ClpA and ClpX likely employ similar “pulling” mechanisms for unfolding and translocation controlled by ATP-driven loop movements within the pore.

I.G. Structural elements important in communication between ClpA/X and ClpP

The degradation of protein substrates by energy-dependent proteases requires communication between the ATPase component (e.g., ClpA or ClpX) and the protease component (e.g., ClpP). Several studies have shown functional communication between these components. The binding of ClpP lowers the ATPase rates of ClpA and ClpX [72, 73], ClpP peptidase activity is stimulated by the binding of ClpA or ClpX [15, 74], and ClpA-ClpP and ClpX-ClpP affinities correlate with both the catalytic status of the ClpP active sites [75] and the nucleotide state of ClpA or ClpX [15, 35, 76]. However, the structural features involved in this communication are incompletely understood. To date, two sets of structural features have been identified which control binding and communication between ClpA/X and ClpP: the IG(F/L) loops on ClpA/X and the N-terminal loops of ClpP.

I.G.1. The IG(F/L) loops of ClpA/X mediate interaction with ClpP.
In 2001, two independent studies showed that the IG(F/L) loops of ClpA and ClpX are important for interaction with ClpP [73, 77]. Subsequent crystal structures showed that these loops are located on the ClpP-binding face of ClpA and ClpX. Although these loops are disordered in the crystal structure of ClpA [16], they are ordered in the ClpX crystal structure and are clearly shown to protrude from the ClpP-binding face of ClpX [17] (green loops in Figure I.6). Furthermore, the IG(F/L) loops appear to be a good predictor of whether or not an ATPase binds ClpP; while these loops are conserved in ClpX and ClpA, they are absent from ClpB [73], a non-ClpP interacting AAA+ molecular chaperone closely related to ClpA in sequence and structure [78]. It has been proposed that the IG(F/L) loops dock into a conserved hydrophobic cleft located on the ATPase-binding face of ClpP [17] (purple surface in Figure I.6). However, due to the symmetry mismatch, each ClpA or ClpX hexamer contains six IG(F/L) loops while ClpP contains seven hydrophobic clefts per heptameric face. How this symmetry mismatch is accommodated is not yet clear. Interestingly, a 2007 study showed that all six IG(F/L) loops are necessary for interaction with ClpP and that this interaction is static; it does not change with the nucleotide state of ClpX [76]. These studies indicate that the IG(F/L) loops may be important in keeping ClpA-ClpP and ClpX-ClpP complexes associated throughout the catalytic cycle.

I.G.2. The N-terminal loops of ClpP mediate interaction with ClpA/X.

The ClpP N-terminal loops, like the IG(F/L) loops of ClpA and ClpX, also mediate ClpA-ClpP and ClpX-ClpP interactions. These loops line the ClpP axial pore and therefore sit at the ATPase-interacting surface of ClpP [26] (Figure I.3 and Figure
Structural elements important in ClpA/X and ClpP binding. A) Top—ClpX IGF loops are shown in green. Pore-2 loops are shown in blue. Bottom—ClpP hydrophobic pockets (in which IG(F/L) loops of ClpA/X bind) are shown in purple. ClpP N-terminal loops are shown in red. B) Model of ClpX-ClpP interaction. The ClpX hexameric model is based on the ClpX monomeric crystal structure [17]. The ClpP structure is from Ref. [29]. The ClpX pore-2 loops are absent from crystal structure and therefore modeled in. Two ClpX subunits are removed to show ClpX pore-2 loop/ClpP N-terminal loop interaction. Color scheme is the same as that of A. Figure adapted from Ref. 76.
In 2007, Martin et al. reported that these loops interact with the pore-2 loops within the central channel of ClpX (Figure I.6) and that this interaction varies with the nucleotide state of ClpX [76]. Additionally, crystal structures of ClpP indicate that the ClpP N-terminal loops are flexible [26-33]. In several structures, all or part of these loops are disordered; however, in structures with ordered loops, the loops assume one of two conformations—a “down” conformation in which the loops are buried in the pore region of ClpP [29], or an “up” conformation in which the loops protrude from the ClpP axial pore [27-30]. In 2006, Bewley et al. reported a crystal structure of ClpP showing six of the seven N-terminal loops in an “up” conformation while one was in a “down” conformation. They hypothesized that this arrangement of the N-termini could induce a symmetry match between ClpP and ClpA/X that might be functionally important [29]. Collectively, these data indicate that the ClpP N-terminus plays a role in communication between ClpA/X and ClpP, and conformational dynamics of this loop region may facilitate this communication.

I.G.3. ATPase-induced gate opening stimulates proteolysis

The flexibility of the ClpP N-terminus, along with the importance of this region in interaction with partner ATPases, suggests that the ClpP N-terminus could gate access to the ClpP proteolytic chamber. Such a gating mechanism is utilized by both the proteasome and HslUV. The axial entry pores of both the 20S proteasome and HslUV are in a closed conformation in the absence of partner ATPases. It has been shown through a
combination of biochemical and structural studies that these pores “open” upon the binding of ATP-bound partner ATPases. For both HslUV and the proteasome, this opening stimulates the peptidase activity of the enzyme by increasing substrate access to the degradation chamber [70, 79-83]. As shown in Chapter 2 of this thesis, ClpAP employs a similar mechanism.

I.H. Processivity of ClpAP

Another defining feature of energy-dependent proteases is processivity. In contrast to extracellular proteases such as trypsin and chymotrypsin, which release intermediates after each peptide bond hydrolysis event, energy-dependent proteases degrade full-length proteins into small peptide products without the release of large intermediates [84]. Processivity requires a driving force; however, this driving force is not necessarily ATP hydrolysis. For example, the 20S proteasome has been shown to degrade full-length proteins processively in the absence of any partner ATPases [85]. However, it has been reported that ClpAP requires the ATP-hydrolysis activity of ClpA to drive processive proteolysis of full-length protein substrates [86]. In Chapter 3 of this thesis, we show that, contrary to previous reports, ClpP, like the 20S proteasome, can processively degrade a full-length protein substrate in the absence of any partner ATPases, albeit it at a greatly reduced rate.

The mechanism of ATP-independent processive proteolysis is not yet clear. One possibility is that peptide bond hydrolysis provides the driving force for processivity. This reaction is exergonic [87] and could drive processive proteolysis through either a power stroke or a ratchet mechanism. Peptide bond hydrolysis was shown previously to
drive processive movement of activated collagenase (MMP-1) on collagen fibrils. In this system, proteolysis fuels biased diffusion of the collagenase enzyme on collagen fibrils via a Brownian-ratchet-like mechanism [88]. In Chapter 3 of this thesis, we propose a similar proteolysis-driven mechanism that accounts for the ATP-independent processivity observed with both ClpAP and the 20S proteasome.

I.I. Control of product sizes by ClpAP

The proteasome, Lon, HslUV, and ClpAP have all been shown to generate peptide products with a size range of 3-30 amino acids [89-95]. Furthermore, the size distribution of peptide products has been measured for both the proteasome and ClpAP; both enzymes have been shown to have an approximately log-normal peptide product size distribution with a peak at 6-9 residues [89, 90, 92]. These data suggest that all energy-dependent proteases share a common mechanism of product size control. However, this mechanism has been disputed. It was originally proposed by Wenzel et al. that peptide product sizes were controlled by a “molecular ruler” mechanism in which the coordinated action of adjacent active sites determines peptide product sizes [96]. According to this mechanism, the spacing between active sites determines peptide product sizes. Indeed, the active sites in both the archaeal 20S proteasome [97] and ClpP [26] are ~25Å apart, equivalent to a peptide length of 7-8 amino acids and consistent with the major peak observed in both the proteasome and ClpAP peptide product size distributions. However, a number of observations are inconsistent with this mechanism. First, although both the proteasome and ClpAP peptide product size distributions have a peak around 7-8 amino acids, a significant portion of the peptide products generated by
these enzymes are longer or shorter than 7-8 residues [89, 90, 92]. Second, although the
degree product size distributions of the archael and eukaryotic proteasomes are
essentially identical [89, 90], the active sites of the 20S archael proteasome are spaced
~25Å apart [97], while the eukaryotic proteasome active sites are irregularly spaced [98].
This observation suggests that the peptide product sizes are independent of the spacing
between active sites. Finally, both the proteasome and ClpAP peptide product size
distributions are unaltered by partial inactivation of the active sites [90, 92], in contrast to
the molecular ruler mechanism which predicts that partial inactivation will increase the
average product size.

As an alternative to the molecular ruler mechanism, both the Goldberg group [90]
and the Flanagan group [34] have proposed a filter mechanism of product size control in
which peptide products are cleaved within the degradation chamber until they are small
enough to diffuse out. However, there are inconsistencies between this mechanism and
the observed properties of the proteasome and ClpAP peptide product size distributions
as well. The filter mechanism implies that peptide product sizes will depend on the
relative rates of proteolysis and diffusion. According to this mechanism, a decrease in
the proteolysis rate should lead to an increase in partitioning towards diffusion out the
chamber, and therefore longer product sizes. However, this is not observed. Partial
inactivation of either the 20S proteasome or ClpP active sites does not alter either the
proteasome or ClpAP peptide product size distributions [90, 92].

In an attempt to accommodate all of the existing experimental data, our lab has
proposed a mechanism of product size control in which peptide product sizes are
controlled by alternation between translocation and proteolysis. In such a mechanism,
peptide product sizes are independent of the relative rates of translocation and proteolysis, consistent with experimental data [92]. The mechanism of peptide product size control is expanded upon in Chapter 3 of this thesis.

I.J. Contents of the thesis

The research presented in this thesis is divided into three chapters:

1) **Chapter 2** – The ClpP N-terminus coordinates substrate access with protease active site reactivity

2) **Chapter 3** – ClpP hydrolyzes a protein substrate processively and independently of the ClpA ATPase: Mechanistic studies of ATP-independent processive proteolysis

3) **Chapter 4** – Synchrotron protein footprinting supports substrate translocation by ClpA via ATP-induced movements of the D2 loop

Chapter 2 addresses the role of the ClpP N-termini in degradation by ClpAP. Synchrotron hydroxyl radical footprinting [99, 100] is used to show that the N-termini of ClpP assume an open, more solvent-exposed conformation upon the binding of ATP-bound ClpA. With additional data from kinetic studies and the use of a ClpPΔN mutant, it is shown that opening of the ClpP pore accelerates the degradation of large peptide substrates by increasing access to the degradation chamber. This chapter also describes surprising data showing that a ClpPΔN truncation mutant is deficient in acyl-enzyme breakdown. These results indicate that the ClpP N-terminal region gates substrate access to the proteolytic chamber, as previously proposed, and suggest a novel role for the ClpP N-termini in controlling acyl-enzyme reactivity at the ClpP active sites. Collectively, the work in Chapter 2 indicates that the ClpP N-termini open to allow substrate entry into the
chamber and acyl-enzyme formation, and that a subsequent conformational change in this region is necessary to allow efficient acyl-enzyme hydrolysis.

Chapter 3 of this thesis addresses the mechanism of processivity and product size control by ClpAP. In this chapter it is shown that ClpP alone can processively degrade a full-length protein substrate, without the assistance of an ATPase component. Furthermore, the size distribution of ClpP-generated peptide products is remarkably similar to the size distribution of ClpAP-generated peptide products, showing that peptide product sizes are largely controlled by ClpP. The size distribution of both ClpAP- and ClpP-generated peptide products can be well-fitted by a sum of three Gaussians with mean values spaced 7-8 amino acids apart. This size distribution is most consistent with a mechanism in which ClpP translocates protein substrates in discrete steps of 7-8 amino acids, and after each step the enzyme partitions between taking another translocation step and proteolysis. It is speculated that the physical basis of this size distribution is the spacing between ClpP active sites, and that processive proteolysis by ClpP is controlled by a Brownian-ratchet-like mechanism in which peptide bond hydrolysis is coupled to the translocation of protein substrates along the active site binding groove of ClpP.

Finally, Chapter 4 of this thesis examines nucleotide-dependent conformational changes in ClpA using synchrotron hydroxyl radical footprinting [99, 100]. Specifically, footprinting is used to test the accuracy of two ClpA hexameric models, both of which are derived from an ADP-bound ClpA monomeric crystal structure [16, 21]. The solvent accessibility results derived from footprinting experiments using ATPγS-induced ClpA hexamers largely agree with the solvent accessibility predictions of the models; however, strong disagreement exists in two regions—the D1 sensor loop and the D2 loop. Both of
these loops lie within the central channel of ClpA. The results are consistent with a nucleotide-dependent conformational change in the ClpA D2 loop. It is proposed that ATP binding causes the D2 loop to assume an “up” conformation in which it contacts the D1 sensor. ATP hydrolysis then induces a conformational change of this loop to the “down” position; this conformation is approximated in the models derived from ADP-bound ClpA. The data presented provide the first structural evidence for a D2 loop conformational change previously proposed to drive substrate translocation through the ClpA pore [21].
I.K. References


Chapter II: The ClpP N-terminus coordinates substrate access with protease active site reactivity

*The synchrotron hydroxyl radical footprinting work described in this chapter was performed in collaboration with Dr. Jen Bohon and Prof. Mark Chance of Case Western Reserve University.

II.A. Summary

Energy-dependent protein degradation machines, such as the E. coli protease ClpAP, require regulated interactions between the ATPase component (ClpA) and the protease component (ClpP) for function. Recent studies indicate that the ClpP N-terminus is essential in these interactions, yet the dynamics of this region remain unclear. Here, we use synchrotron hydroxyl radical footprinting and kinetic studies to characterize functionally important conformational changes of the ClpP N-terminus. Footprinting experiments show that the ClpP N-terminus becomes more solvent-exposed upon interaction with ClpA. In the absence of ClpA, deletion of the ClpP N-terminus increases the degradation rate of large peptide substrates 5-15 fold. Unlike ClpAP, ClpPΔN exhibits a distinct slow phase of product formation that is eliminated by the addition of hydroxylamine, suggesting that truncation of the N-terminus leads to stabilization of the acyl-enzyme intermediate. These results indicate that 1) the ClpP N-terminus acts as a “gate” controlling substrate access to the active sites, 2) binding of ClpA opens this “gate”, allowing substrate entry and formation of the acyl-enzyme intermediate, and 3) closing of the N-terminal “gate” stimulates acyl-enzyme hydrolysis.

Note: A manuscript describing this work entitled “Jennings, L.D., Bohon, J., Chance, M.R., & Licht, S. The ClpP N-terminus coordinates substrate access with protease active site reactivity” has been accepted for publication in Biochemistry.
II.B. Introduction

Energy-dependent protein degradation plays a vital role in protein homeostasis in all organisms [1]. Examples of proteolytic machines include the eukaryotic 26S proteasome and the bacterial proteases, FtsH, Lon, HslUV, ClpXP, and ClpAP. In these proteolytic complexes, AAA+ ATPases (ATPases associated with various cellular activities) recognize, unfold, and translocate protein substrates into barrel-shaped compartmental proteases where the proteins are hydrolyzed and released as peptide products [2-4]. Despite significant sequence divergence, the overall architecture of these complexes is conserved. For example, in ClpAP, ClpXP, HslUV, and the 26S proteasome, the protease subunits (ClpP, HslV, or the 20S peptidase) form either hexameric or heptameric ring structures that stack face to face, resulting in the sequestering of the active sites in a solvent-exposed chamber [5-8]. The ATPase subunits (ClpA, ClpX, HslU, or the 19S complex) form hexameric ring-shaped structures that stack co-axially with the ring-shaped proteases, creating a continuous pore through which denatured protein substrates can be passed from the ATPase subunit into the protease compartment [9-13].

Understanding the communication between the ATPase component and the protease component is a key element in understanding the mechanism of protein degradation by these machines. A number of domain interactions have been shown to participate in communication between the protease ClpP and its partner ATPases, ClpX and ClpA. The IGF/L loops of ClpA and ClpX that interact with ClpP are important in the formation of ClpA/XP complexes [14-16], with all six loops being required for complex formation [17]. In addition, the N-terminal ClpP axial loops interact with the pore-2 loops of ClpX [17], and mutations and deletions in the ClpP N-terminal region
abolish binding to both ClpX and ClpA [18-20].

Although these previous studies revealed important interactions between ClpP and ClpX/ClpA, it remains an open question as to how the IGF/L loops and the N-terminal ClpP loops may change conformation in the course of the catalytic cycle. The sequence of the ClpP N-terminus is highly conserved across many species [18, 19]; however, structural studies indicate that that this region of ClpP is flexible [5, 18-23]. These observations suggest that this region may play an important dynamic role in the processing of peptide substrates. ClpP structures have been solved from five different organisms, and multiple crystal structures exist for some of these species [5, 18-23] (see Ref. [24] for a review). While the overall structure of ClpP is highly conserved among the structures solved to date, a high degree of variability exists in the conformation of the N-terminal region. In structures for which electron density is observable, the N-terminus crystallizes in one of two ways—in an “up” conformation or in a “down” conformation. In the “up” conformation, which has been observed in *Escherichia coli* [20, 21], *Streptococcus pneumoniae* [19], and human [18] enzymes, the N-terminus forms a loop in which the first ~7 residues line the axial pore while residues ~8-16 form a flexible loop extending out of the pore. The “down” conformation (observed in *Escherichia coli* [20]) is less well-defined. In this form, no residues are visibly protruding from the pore; electron density assigned to the first eleven residues is present in the pore, but is poorly defined. In their 2006 paper, Bewley *et al.* published a ClpP structure in which 6 of the N-terminal loops on a ClpP heptamer were in the “up” conformation while one was in the “down” conformation [20]. They suggested that this arrangement of N-termini may facilitate a local symmetry match between ClpP and ClpX/ClpA, which may play a
critical role in complex formation and/or catalysis.

The flexibility of the ClpP N-terminus exhibited in multiple crystal structures, along with the recent finding that it is important in interactions with its partner ATPases, suggests that the ClpP N-terminus could gate access to the ClpP degradation chamber. Such a mechanism is utilized in the proteasome. The N-termini of the 20S core particle alpha subunits occupy the pore region of this cylindrical protease. Access to the active sites is permitted only by removal of these loops, which is facilitated by the binding of its partner ATPase, the 19S regulatory particle [25-29]. A similar mechanism may be employed by ClpAP and ClpXP. When exit of peptide products through the equatorial pores of ClpP is blocked, truncation of the N-terminus allows these products to escape the ClpP complex [30], suggesting that occupancy of the N-terminus in the ClpP pore interferes with transport of peptides through the complex.

Here, we describe synchrotron hydroxyl radical footprinting experiments that identify conformational changes that occur in ClpP upon binding to ClpA. This technique permits relative quantification of the solvent accessibility of protein side chains, allowing interaction sites and conformational changes to be mapped to specific areas of the protein. Synchrotron protein footprinting requires the generation of a transient concentration of hydroxyl radicals by direct irradiation of the protein solution for milliseconds with a high-flux x-ray beam. These radicals produce characteristic oxidative modifications on protein side chains [31, 32] in solvent-exposed regions of the protein; areas of the protein not exposed to solvent are protected from these stable oxidations. Protease digestion of the sample followed by liquid-chromatography-coupled mass spectrometry (LCMS) and tandem mass spectrometry (MSMS) allows identification
of the oxidized regions and often the specific residue that is modified as well. Binding sites and regions involved in conformational change are then identified by differences in side chain modification rate upon addition of binding partners or small molecule ligands [33, 34].

The synchrotron hydroxyl radical footprinting results show that the ClpP N-terminal residues are shielded from solvent when ClpP is not in complex with ClpA but become significantly more solvent-exposed upon binding of ClpA. To investigate the functional consequences of this conformational change, we also carried out steady-state kinetic characterization of wild-type and N-terminally truncated [19] ClpP. These kinetic studies reveal that removal of the first seven residues of ClpP increases peptidase activity by 5-15 fold. These results provide the first direct evidence that the ClpP N-terminus controls the enzyme’s catalytic activity. Moreover, we show here that removal of the N-terminus traps the enzyme in an acyl-enzyme bound form, indicating that a conformational change involving the N-terminus is necessary for acyl-enzyme intermediate breakdown. Together, these results support the hypothesis that motions of the ClpP N-terminus in and out of the ClpP pore regulate both substrate access to the protease active sites and the catalytic efficiency of these active sites for acyl-enzyme hydrolysis.

II.C. Results

Synchrotron hydroxyl radical footprinting

To probe the conformational dynamics of ClpP upon binding of ClpA, we performed synchrotron hydroxyl radical footprinting analysis of the ClpP tetradecamer
alone and with ClpA bound. MSMS-verifiable coverage of the ClpP tryptic digest encompassed 49% of the protein (Figure II.1.A). Only one tryptic peptide (the N-terminal peptide, residues 1-12) exhibited modification upon x-ray exposure; this low level of observable modification is likely due to the presence of 1 mM ATPγS, a strong hydroxyl radical scavenger that is required for formation of the ClpA hexamer. Nevertheless, the difference in modification rate of this region for the ClpP tetradecamer vs. the ClpAP complex is significant and reveals an important ClpA-induced conformational change in ClpP. The most abundant peptide modification observed was a +16 Da mass shift on methionine 5 (Figure II.2), accompanied by a substantial decrease in retention time (~2 minutes) during chromatographic separation (Figure II.3). A +14 Da mass shift on the N-terminal alanine was also observed, but did not contribute significantly to the computed modification rate. The modification rate of this region for ClpP alone is 1.5±0.5 s⁻¹ (the modification rates reported are x-ray dependent rates; the fraction of peptide modified before x-ray exposure is subtracted in the kinetic analysis). Formation of the ClpAP complex might be expected to decrease this rate due to obstruction of free solvent access to the ClpP tetradecamer by the proximity of bound hexameric ClpA. In fact, binding of ATPγS-bound ClpA increases the rate by almost 6-fold to 8.5±1.7 s⁻¹ (Figure II.1.B), indicating that the binding of ClpA causes the ClpP N-termini to assume a more solvent-exposed conformation. The relatively low modification rate observed for ClpP alone suggests that the ClpP N-termini may be primarily in a “closed”, pore-filled conformation.
Synchrotron footprinting tandem MS spectra. Shown are spectra of the unmodified (top), modified methionine 5 (middle) and modified alanine 1 (bottom) ClpP N-terminal (1-12) peptides. Observed y- and b-ions are labeled.
Synchrotron footprinting chromatographic traces. Shown are traces of the unmodified (top), modified methionine 5 (middle) and modified alanine 1 (bottom) ClpP N-terminal (1-12) peptides.
when not bound to ClpA; the high modification rate observed for ClpAP may indicate an “opening” of the ClpP pore through removal of the N-termini upon the binding of ClpA. 

*The ClpPΔN mutation accelerates degradation of insulin chain B and induces distinct rapid and slow phases of product formation.*

Our hydroxyl radical footprinting data suggested an “opening” of the ClpP N-termini upon ClpA binding; in order to test this hypothesis, we turned to kinetic studies. To mimic the ClpP open conformation, we expressed and purified a truncation mutant (ClpPΔN) in which the first seven residues of the mature N-terminus are removed [19]. Previous studies [18, 19] have shown that this deletion greatly diminishes binding to both ClpX and ClpA. Therefore, all studies with this mutant were performed in the absence of either ATPase.

Previous work has demonstrated that wild-type ClpP is inefficient at degrading large (~30 amino acid) peptides such as oxidized insulin chain B and glucagon. However, this work also showed that the addition of ClpA and either ATP or the nonhydrolyzable analog ATPγS greatly increased the efficiency of degradation [35, 36]. Based on our hydroxyl radical footprinting data, we hypothesized that the increase in degradation efficiency induced by the ClpA hexamer was due to opening of the ClpP pore by removal of the N-terminus. This hypothesis predicts that the ClpPΔN mutant will mimic ClpAP in the processing of large substrates. To test this hypothesis, we compared the insulin chain B degradation kinetics of ClpPΔN to those of ClpAP and wild-type ClpP.

Degradation of the 30-amino-acid-long insulin chain B yields four major products and several less-abundant products (Figure II.4.A). ClpPΔN, ClpAP, and wild-type ClpP
**Insulin chain B degradation.** A) HPLC trace of insulin chain B degradation. Shown in black is degradation of 50 μM insulin chain B with 1 μM ClpP_14_, 2 μM ClpA_6_, and 2 mM ATPγS at t=0 min. Shown in red is t=10 min. The most abundant product is labeled with an asterisk. All data shown in B-D are with the most abundant product. B) Michaelis-Menten plots with ClpAP, wild-type ClpP, and ClpPΔN. ClpPΔN rapid phase rates shown in red and slow phase rates shown in orange. Values for $K_m$ are 350±60 μM, 340±80 μM, and 330±90 μM for wild-type ClpP, ClpPΔN, and ClpAP, respectively. Values for apparent $k_{cat}$ are 2.4±0.2 min⁻¹, 13±1 min⁻¹, and 19±2 min⁻¹ for wild-type ClpP, ClpPΔN, and ClpAP, respectively. C) ClpAP, wild-type ClpP, and ClpPΔN degradation of 10 μM insulin chain B with 1 μM enzyme. Only ClpPΔN shows a rapid phase. D) ClpPΔN degradation of insulin chain B at various concentrations of substrate. Only low substrate concentrations exhibit a distinct rapid phase. Error bars represent standard deviation of three trials.
produces the same four major products, as judged by HPLC retention time. The formation rates of all four products were measured for each of the three enzyme conditions, and the steady-state kinetic parameters were determined. Plots of the formation rate of the most abundant product as a function of concentration are shown in Figure II.4.B. Wild-type ClpP, ClpAP, and ClpPΔN each exhibit a $K_m$ for insulin chain B of $\sim$350 µM. However, ClpAP and ClpPΔN exhibit significantly higher maximal formation rates for each of the major peptide products compared to wild-type ClpP. The apparent $k_{cat}$s for each of the four major peptides produced by ClpAP are 8-20 fold higher than those measured using wild-type ClpP, while the maximal formation rates with ClpPΔN are 5-15 fold higher than wild-type ClpP (Figure II.4.B).

Unexpectedly, and unlike wild-type ClpP and ClpAP, at low concentrations of insulin chain B (≤150 µM) the rate of product formation by ClpPΔN exhibits a rapid phase followed by a slow phase in which product formation was essentially undetectable (Figure II.4.C). The amplitude of this rapid phase depends on the substrate concentration, with higher amplitudes observed at higher substrate concentrations (Figure II.4.D). At insulin chain B concentrations of 250 µM and higher, only the rapid phase is evident (Figure II.4.D). Furthermore, the largest rapid phase amplitude observed (at 150 µM insulin chain B) is approximately stoichiometric with the active site concentration (i.e., 1 equivalent of product was generated for each equivalent of protease active site in the ClpP tetradecameric complex; in other words, $\sim$14 equivalents of product were generated per ClpPΔN$_{14}$ tetradecamer), indicating that the rapid phase represents at least one complete substrate turnover. The observation of a distinct slow phase in ClpPΔN kinetics suggests that a step that is rapid for the wild-type enzyme becomes rate-limiting in the
absence of the ClpP N-terminus; one possibility (precedented in serine proteases, especially with ester substrates [37]) is that acyl-enzyme breakdown is rate-limiting. The data also indicate that the substrate concentration affects the ability of the enzyme to process and/or bypass the long-lived species that predominates in the slow phase of product formation.

**Kinetic contributions of substrate entry, acyl-enzyme formation, and acyl-enzyme breakdown**

Comparing the kinetics of large and small peptide substrates is one way to determine the relative contributions of substrate entry and subsequent covalent chemical steps to the rapid and slow phases of product formation. Previous work has demonstrated that addition of ClpA and ATP or ATPγS slightly decreases the degradation rate of the small peptide substrate, N-succinyl-Leu-Tyr-7-amido-4-methyl-coumarin (SLY-AMC) compared to the rate seen with ClpP alone (~3-fold decrease in rate at a concentration of 1 mM) [35]. Previous results have also shown that truncation of the ClpP N-terminus has a relatively modest effect on the degradation rate of SLY-AMC (~2-fold increase in the steady-state rate at a concentration of 0.5 mM) [19], consistent with the hypothesis that this substrate can enter either through the axial or equatorial pores of ClpP. These results suggest that access to the ClpP degradation chamber is not rate-limiting for small substrates such as SLY-AMC. If the rapid phase of the ClpPΔN-catalyzed reaction corresponds to substrate entry and acyl-enzyme formation, while the slow phase corresponds to the breakdown of a stabilized acyl-enzyme, rapid and slow phases will be observed for ClpPΔN-catalyzed proteolysis of the small substrate. However, if access to the ClpP active sites is not rate-limiting for small substrates, removal of the N terminus
will not accelerate the rapid phase compared to the wild-type rate. In order to test these hypotheses, we measured the degradation kinetics of SLY-AMC with wild-type ClpP, ClpPΔN, and ClpAP.

The previously reported $K_m$ of SLY-AMC with wild-type ClpP is 1 mM [38]. However, we observe an apparent decrease in the steady-state rate of degradation at SLY-AMC concentrations greater than 250 μM. We suspect that this decrease in apparent rate is due to the inner filter effect (absorption of emitted photons by dye molecules in highly absorbing solutions [39]), and therefore our rates above 250 μM SLY-AMC are not accurate. Nevertheless, we were able to compare the degradation rates of SLY-AMC with wild-type ClpP, ClpAP, and ClpPΔN at various concentrations of substrate. At all concentrations used, the initial degradation rate is the same within error for wild-type ClpP, ClpAP, and ClpPΔN (Figure II.5.A). (Note: Previous studies have reported that ClpPΔN hydrolyzes SLY-AMC 2-fold more rapidly than the wild-type enzyme, based on rates at a single substrate concentration (0.5 mM) [19]. We also observe somewhat more rapid rates for ClpPΔN at high substrate concentrations (Figure II.5.A). However, at these concentrations, the apparent rate decreases with increasing substrate concentration due to substrate inhibition and/or inner-filter effects. We have therefore chosen to use the maximal hydrolytic rate constant in interpreting this kinetic data.) The lack of rate enhancement seen with ClpPΔN as compared to wild-type ClpP supports the hypothesis that the rapid kinetic phase corresponds to substrate entry and/or acyl-enzyme formation. It also suggests that, in contrast to the observation for large substrates, removal of the N-terminal region of ClpP does not accelerate substrate entry/acyl-enzyme formation for small substrates.
Degradation of SLY-AMC. A) ClpAP, wild-type ClpP, and ClpPΔN degradation rate constants as a function of SLY-AMC concentration. B) Kinetics of 50 μM SLY-AMC degradation with ClpAP, wild-type ClpP, and ClpPΔN (using 0.125 μM ClpP14). Only ClpPΔN exhibits a distinct rapid phase. C) ClpPΔN degradation of SLY-AMC at various concentrations of substrate. A rapid phase is present at concentrations ≤250 μM. The apparent change in rate seen with 500 μM and 1 mM is also present with wild-type ClpP and ClpAP and is due to saturation of fluorescence signal; it does not represent a distinct rapid phase. Traces shown in B and C are the average of 3 trials.
As observed with insulin chain B, the kinetics of SLY-AMC degradation by ClpPΔN exhibit a rapid phase comprising multiple turnovers followed by a slow phase, whereas no slow phase is seen with wild-type ClpP or ClpAP (Figure II.5.B). Consistent with the insulin chain B results, distinct rapid and slow phases are only observed at low (≤250 μM) concentrations of SLY-AMC, and the amplitude of the ClpPΔN rapid phase increases with increasing substrate concentration (Figure II.5.C). In this case, the largest rapid phase amplitude observed (at a SLY-AMC concentration of 250 μM) is greater than stoichiometric: 5 equivalents of product are generated per equivalent of protease active site in the complex (in other words, ~70 equivalents of product are generated per ClpPΔN_{14} tetradecamer), indicating that the rapid phase represents multiple substrate turnovers. These observations show that increased substrate concentrations help the enzyme achieve catalytic turnover rather than becoming trapped in a stabilized intermediate state.

*Hydroxylamine eliminates the ClpPΔN slow phase*

The existence of a slow phase in the ClpPΔN degradation kinetics suggests that a step following acyl-enzyme formation is rate-limiting for catalysis at low substrate concentrations. One possibility is that breakdown of a stabilized acyl-enzyme intermediate is the rate-limiting step. In order to test this hypothesis, we measured the kinetics of SLY-AMC degradation in increasing concentrations of the strong nucleophile hydroxylamine [40-42] (Figure II.6.A). We found that the slow-phase rate of ClpPΔN degradation increases with increasing concentrations of hydroxylamine, and at 2.4 M NH₂OH the slow phase is not observed (Figure II.6.A and II.6.B). These data are
The effects of hydroxylamine on the degradation of SLY-AMC. **A)** Time course of ClpPΔN degradation of SLY-AMC. **B)** Rate constant of SLY-AMC degradation by ClpPΔN as a function of hydroxylamine concentration. **C)** Time course of wild-type ClpP degradation of SLY-AMC. **D)** Rate constant of SLY-AMC degradation by wild-type ClpP as a function of hydroxylamine concentration. Reaction conditions: 50 μM SLY-AMC with 0.125 μM ClpP tetradecamer. Traces shown are the average of three trials, and error bars represent the standard deviation of three trials.
consistent with hydroxylamine increasing the ClpPΔN slow-phase rate of degradation by accelerating the rate of acyl-enzyme breakdown.

In contrast, the rate of degradation by wild-type ClpP decreases with increasing concentrations of hydroxylamine, as does the rapid phase rate of degradation with ClpPΔN (Figure II.6.B, II.6.C, II.6.D). This inhibitory effect of high (~molar) concentrations of hydroxylamine on the rapid phase of turnover does not have an obvious specific molecular mechanism. However, the effects of hydroxylamine on the rapid phase are similar for wild-type ClpP and ClpPΔN and were therefore not interpreted further in terms of possible functions for the N-terminal domain. The possibility that hydroxylamine has non-specific effects on the rapid kinetic phase nonetheless highlighted the importance of additional experiments to determine whether its effects on the slow phase are due to acceleration of acyl-enzyme breakdown.

*Hydroxylamine traps a stabilized acyl-enzyme intermediate*

In order to more rigorously test the mechanism of ClpPΔN slow-phase rate acceleration by hydroxylamine, we used MALDI-MS analysis to test for the presence of hydroxamic acid products following addition of hydroxylamine. However, the low molecular weight of the peptide products of SLY-AMC degradation precludes their observation using MALDI-MS. The complex product mixture obtained upon degradation of insulin chain B (Figure II.4.A) also complicates interpretation of the MALDI spectrum. To simplify the analysis, we used as a substrate the ssrA degradation tag (AANDENYLAA), derivatized with a dabsyl group at the N-terminus to facilitate detection by UV absorbance. Previous studies in our lab indicated that degradation of this substrate yields two primary products (AANDENYA + LAA) (data not shown).
First, we characterized the degradation of ssrA-dabsyl with wild-type ClpP, ClpAP/ATPγS, and ClpPΔN. At 10 μM ssrA-dabsyl, we found that the initial rate of degradation of this intermediate-sized substrate is 3-fold faster with ClpAP as compared to wild-type ClpP and 5-fold faster with ClpPΔN as compared to wild-type ClpP (Figure II.7). This is consistent with the insulin chain B data, and indicates that removal of the N-terminus, either through truncation or by complex with ClpA, increases the rate of acyl-enzyme formation for this 11-amino acid long substrate. As expected, we also observed distinct rapid and slow phases in the kinetics of ClpPΔN degradation (Figure II.7).

Upon addition of hydroxylamine, we observed the formation of a new peak in the HPLC trace (Figure II.8.A). MALDI analysis confirmed that this new peak is the hydroxamic acid peptide product (Figure II.8.B). Kinetic analysis revealed that, as observed with SLY-AMC degradation, the addition of increasing concentrations of hydroxylamine accelerates the slow-phase rate of ClpPΔN degradation (Figure II.9A and II.9.B). This is achieved largely through formation of the hydroxamic acid product, although we did see a modest increase in the production of the carboxylic acid product. (At 1.6 M NH₂OH the concentration of the carboxylic acid product increased ~2-fold, Figure II.8.A.) As a control, we measured the kinetics of degradation with wild-type ClpP at increasing hydroxylamine concentrations. We found that while addition of hydroxylamine does cause limited formation of the hydroxamic acid product, the rates of formation are significantly slower than the rates of formation with ClpPΔN. For example, the rate of hydroxamic acid production is 15-fold slower at 1.6 M NH₂OH with wild-type ClpP as compared to ClpPΔN (Figure II.8.D). As observed with SLY-AMC,
**Degradation of ssrA-dabsyl.** Conditions: 10 μM ssrA-dabsyl with 0.125 μM enzyme. Rate constants are 0.53 min⁻¹, 1.42 min⁻¹, 2.51 min⁻¹, and 0.02 min⁻¹ for wild-type ClpP, ClpAP/ATPγS, ClpPΔN rapid phase rate and ClpPΔN slow phase rate, respectively. Error bars represent the standard deviation of two trials.
Degradation of ssrA-dabsyl with hydroxylamine produces the hydroxamic acid product. A) ClpΔN degradation of ssrA-dabsyl (0.125 μM ClpΔN_{14}, 10 μM ssrA-dabsyl). Shown in black is the reverse-phase chromatogram of the reaction mixture after 30 min. in the absence of NH$_2$OH. Shown in red is the chromatogram of the reaction mixture after 30 min. in the presence of 1.6 M NH$_2$OH. B) Concentration of the hydroxamic acid product over time with 1.6 M NH$_2$OH. (conditions: 0.125 μM ClpP tetradecamer, 10 μM ssrA-dabsyl). Error bars represent standard deviation of two trials. C) MALDI spectrum of carboxylic acid product peak. D) MALDI spectrum of hydroxamic acid product peak.
The effects of hydroxylamine on the kinetics ssrA-dabsyl degradation. A) Time course of ClpPΔN degradation of ssrA-dabsyl. B) Rate constant of ssrA-dabsyl degradation by ClpPΔN as a function of hydroxylamine concentration. C) Time course of wild-type ClpP degradation of ssrA-dabsyl. D) Rate constant of ssrA-dabsyl degradation by wild-type ClpP as a function of hydroxylamine concentration. Reaction conditions: 10 μM ssrA-dabsyl with 0.125 μM ClpP tetradecamer. Traces shown are the average of two trials, and error bars represent the standard deviation of two trials. For C and D rates are shown as a fraction of the initial rate at 0 M NH₂OH.
the wild-type ClpP degradation rate decreases with increasing concentrations of hydroxylamine, as does the rapid phase rate of degradation by ClpPΔN (see Figure II.9.B, II.9.C, II.9.D). These data show that hydroxylamine increases the ClpPΔN rate of acyl-enzyme breakdown through formation of the hydroxamic acid product, consistent with the trapping of an acyl-enzyme intermediate.

**II.D. Discussion**

Recent biochemical studies have shown that the N-terminus of ClpP is an important component in the interaction with its partner ATPases ClpX and ClpA [17-20]. Furthermore, the conformational flexibility of the N-terminus has been highlighted in a series of ClpP structures showing the N-terminus in both an “up” conformation in which the N-terminal residues form a well-ordered β-sheet outside the central pore and a “down” conformation in which the first several N-terminal residues are disordered and appear to occupy the central pore [18-21]. This conformational flexibility suggests that while the N-terminus is important for interaction with ClpX and ClpA, it may also be playing a more dynamic role in the overall processing of peptide and protein substrates.

Synchrotron hydroxyl radical footprinting was used to probe the dynamics of the ClpP N-terminal region. These experiments show that the first few residues of the ClpP N-terminus become more solvent-exposed upon interaction with ATPγS-bound ClpA. This observation is consistent with ClpP changing from a “closed” conformation to an “open” conformation upon binding to ClpA. Structurally, this may correspond to a change from the “down” conformation to the “up” conformation [20]. The previous observations that the ClpP N-terminus is required for ClpA/ClpP interaction [19, 20] and
that the N-terminus can functionally occlude the axial pore of the complex [30] suggested
that the N-terminus exits the ClpP pore and resides at the ClpA/ClpP interface. The
footprinting results provide direct structural evidence for this hypothesis in the intact
ClpAP complex.

The relatively low coverage of ClpP (49%) precludes a definitive statement that
domain motions outside of the N-terminus do not occur. However, while the coverage is
incomplete, it spans most of the ClpP sequence; the only large regions of the protein that
are not covered at all are sheets 3-6 and helix C (as defined in ref. 5). It is therefore
unlikely that a global conformational change involving large changes in solvent exposure
has been missed due to limited peptide coverage.

Complementary kinetic experiments provide evidence that this “opening” of the
N-terminus is functionally important in peptide processing. Previous studies have
reported that the processing of long (~30 amino acid) peptides by ClpP is significantly
accelerated by the addition of ClpA and ATPγS. The current experiments indicate that an
N-terminal ClpP truncation mutant mimics the ClpAP state by accelerating the
degradation rate of a 30-amino-acid-long peptide (insulin chain B) 5-15-fold above wild-
type ClpP levels. This rate acceleration was also observed for an 11-amino-acid-long
peptide (ssrA-dabsyl), but not for the short peptide SLY-AMC. These results are
consistent with a model in which the ClpP serine protease active sites are accessible to
large peptide substrates only when the N-terminus is in the “open” conformation, but
small peptides may enter the ClpP chamber through the equatorial pores. The “open”
conformation can be mimicked by an elimination of the first seven N-terminal residues
(ClpPΔN) or by removal of the ClpP N-terminal residues from the pore of ClpP upon
interaction with ATP-bound ClpA.

The unexpected observation of distinct rapid and slow phases for the ClpPΔN mutant kinetics supports a novel role for the ClpP N-terminus in regulating the reactivity of the acyl-enzyme intermediate. Studies with three peptides ranging in size from 2 amino acids to 30 amino acids showed distinct rapid and slow phases in product formation kinetics with ClpPΔN, but not with ClpAP or wild-type ClpP. The N-terminal truncation induces a slow phase without decreasing the rate of the rapid phase, indicating that this structural perturbation has a selective effect on a subset of microscopic reaction steps rather than acting simply to disrupt the enzyme’s structure. The slow phase is eliminated at high concentrations of the strong nucleophile hydroxylamine with concomitant formation of the hydroxamic acid product, consistent with the ClpPΔN mutant trapping the active-site serine in a slowly hydrolyzed acyl-enzyme form that is broken down by hydroxylaminolysis. The slow phase for ClpPΔN can be interpreted as breakdown of the acyl-enzyme intermediate, which appears to be highly stabilized in the absence of the ClpP N-terminus.

The rapid phase in ClpPΔN kinetics is similar to the well-characterized burst phase kinetics of other serine proteases in the sense that formation of the putative acyl-enzyme intermediate occurs during this phase. However, the observation that multiple turnovers can take place during the rapid phase is unusual. The production of multiple equivalents of product followed by a nearly complete loss of activity suggests that active ClpP can undergo a global shift to an inactive conformation that does not hydrolyze the acyl-enzyme intermediate (Figure II.10.A). In that case, the number of turnovers before
Schematic model showing the gating function of the ClpP N-terminus. A) Degradation by ClpPΔN. After formation of the acyl-enzyme intermediate, the enzyme partitions between an active state that is competent for acyl-enzyme breakdown and an inactive state (shown in green) in which the acyl-enzyme intermediate is stabilized. Binding of a second substrate to the putative N-terminal binding site accelerates the rate of entry into the active (hydrolysis competent) state. B) Degradation by ClpAP. ClpA is shown in red, and ClpP is shown in blue. Binding of ClpA causes removal of the ClpP N-terminus from the pore of ClpP, allowing substrates to enter the protease chamber and form the acyl-enzyme intermediate. Re-entry of the N-terminal loops into the ClpP pore causes acyl-enzyme breakdown.
inactivation will depend on the relative rates of acyl-enzyme hydrolysis and entry into the inactive conformation. The observation that the inactive conformation is only present for the ClpPΔN mutant indicates that the N-terminus regulates acyl-enzyme intermediate reactivity by regulating the ability of ClpP to maintain a conformation competent for acyl-enzyme hydrolysis.

Motion of the N-terminus from the “up” to the “down” conformation may provide the specific structural trigger for the conformational change that allows efficient acyl-enzyme hydrolysis in the wild-type enzyme. Szyk et al. have recently reported the crystal structure of wild-type ClpP with an inhibitor peptide covalently bound to the active site, mimicking the tetrahedral intermediate [21]. The structure shows the N-termini of ClpP in the “up” conformation and suggests that the acyl-enzyme intermediate may be stabilized when the ClpP N-termini adopt this conformation. This may also explain why ClpPΔN, which functionally mimics the “open” or “up” conformation, stabilizes the acyl-enzyme intermediate. Isolation of the intact acyl-enzyme intermediate itself would open up many new avenues for understanding its structure and reactivity, especially its remarkable stability at low substrate concentrations. The kinetics of the slow phase suggest that this intermediate will be stable enough to isolate; however, our attempts to do so have thus far been unsuccessful.

The observation that high substrate concentrations increase the amplitude of the ClpPΔN rapid kinetic phase is also consistent with the idea that peptide occupancy within ClpP helps regulate the reactivity of the acyl-enzyme intermediate. One possibility is that occupancy of the axial pores by peptides (either the N-terminal peptide or substrate peptides) triggers the shift to an active conformation. Because ClpP has a large (50 Å)
central cavity, substrate peptides and a translocating polypeptide might simultaneously occupy the axial pore and the central cavity. If in the wild-type enzyme an N-terminal conformational change from the “up” conformation to the “down” conformation (i.e., the “closed”, pore-filled conformation) causes breakdown of the acyl-enzyme intermediate (Figure II.10.B), in the ClpPΔN enzyme a substrate peptide at high concentration might fill the pore in place of the N-terminus, therefore facilitating acyl-enzyme hydrolysis (Figure II.10.A). In this ClpPΔN mechanistic model, increasing the substrate concentration increases the probability that the enzyme hydrolyzes the acyl-enzyme intermediate rather than assuming the hydrolysis-incompetent form. This model thus explains the observation that increasing the substrate concentration increases the rapid phase amplitude and accounts for a rapid phase composed of multiple turnovers. In fact, a simplified quantitative model can be used to illustrate how the kinetic partitioning between the active and inactive conformations can predict the amplitude of the rapid phase (i.e., the average number of turnovers before the enzyme transitions into the inactive conformation).

The probability \( p \) that the enzyme proceeds through the active conformation rather than entering the inactive conformation can be expressed as:

\[
p = \frac{k_{\text{act}} [S]^n}{k_{\text{act}} [S]^n + k_{\text{inact}}},
\]

where \( k_{\text{act}} \) is the rate constant for transition into the active conformation, \( k_{\text{inact}} \) is the rate constant for transition into the inactive conformation, [S] is the concentration of substrate peptide, and \( n \) is the effective kinetic order of substrate peptide binding. This expression allows the dependence of \( p \) on [S] to be determined for arbitrary values of \( k_{\text{act}}, k_{\text{inact}}, \) and
n (Figure II.11.A). With increasing [S], the probability of productive turnover increases, approaching unity in the limit of [S]→∞. The increase of p with increasing [S] becomes steeper when n>1.

The number of turnovers before transition into the inactive conformation will be described by a geometric distribution [43] with expected value (1-q)/q, where q=1-p. This expected value gives the amplitude of the rapid phase. A plot of the dependence of this quantity on [S] for n=1 and n=2 shows that the rapid phase amplitude is proportional to [S]^n (Figure II.11.B). Experimentally, the rapid phase appears to exhibit a steep dependence on [S] (Figure II.4.D and Figure II.5.C); this observation is consistent with substrate and/or N-terminal peptide binding at multiple sites being required for efficient partitioning into the active state. In practice, an absolute requirement for multiple binding events is likely to be an oversimplification, but a similar analysis holds for MWC cooperativity (in that case, n is analogous to the Hill coefficient, and can have non-integer values). Such cooperativity might also be observed if acylation of the active site serine drives the enzyme into its activated conformation.

Our data are most consistent with a mechanism in which increased substrate binding at the putative N-terminal peptide binding sites within the central ClpP pore causes accelerated acyl-enzyme breakdown. Nevertheless, at this point we cannot rule out the possibility that acylation of the active site serine also helps to shift the conformational equilibrium that triggers acyl-enzyme hydrolysis. In that case, high substrate concentrations might favor acyl-enzyme hydrolysis by increasing the steady-state concentration of the acyl-enzyme intermediate. Indeed, the apparent cooperativity observed in partitioning into the active state of the enzyme could be driven by acylation
Dependency of partitioning ratio and rapid phase amplitude on substrate concentration and the kinetic order of substrate binding. A) Dependence of $p$ (probability of partitioning to productive turnover) on [S] (substrate concentration). $k_{act}$ is set at 1, and $k_{inact}$ is set at 0.1 (arbitrary units) for the purposes of illustration. B) Dependence of rapid phase amplitude (average number of turnovers in the rapid phase) on [S] (substrate concentration). $k_{act}$ is set at 1, and $k_{inact}$ is set at 0.1 (arbitrary units) for the purposes of illustration. The rapid phase amplitude is proportional to $[S]^n$, where $n$ is the number of substrate molecules bound in the active form of the enzyme.
of the active site serine rather than peptide binding in the ClpP pore. However, if acylation is the allosteric trigger for acyl-enzyme breakdown, one would expect the wild-type enzyme to also exhibit biphasic kinetics at subsaturating concentrations of substrate; such kinetics are not observed.

Regardless of the exact mechanism(s) at work, the change from the active to the inactive state appears to be an allosteric cooperative transition. Other ring-shaped enzymes, such as the chaperonin GroEL ATPase, couple enzymatic activity to a cooperative allosteric conformational change [44], and the catalytic cycle of ClpP may be analogous.

Taken together, the footprinting and kinetic data suggest a new model for proteolysis by wild-type ClpAP (Figure II.10.B). In this model, an interaction with the ATP-bound form of ClpA causes the ClpP N-terminus to assume the “up” conformation. This conformational change opens the ClpP axial pore, providing substrate access to the ClpP serine protease active sites and enabling formation of the acyl-enzyme intermediate. Re-entry of the ClpP N-terminus into the axial pore (i.e., the “down” conformation) leads to efficient hydrolysis of the acyl-enzyme intermediate and escape of the second peptide product via the equatorial pores. One important unresolved question is exactly how ATP-driven conformational changes of ClpA are coupled to the conformational changes of ClpP that regulate acyl-enzyme reactivity. It may be that motions of the ClpA D2 loop, previously proposed to drive substrate translocation [45], also drive ClpP conformational changes. Further footprinting and kinetic studies are expected to help resolve this question.
II.E. Materials and Methods

Protein purification

C lpP-His$_6$ and ClpA were purified as previously described [38, 46, 47] (see also Appendices A and B). The plasmid encoding ClpPΔN (in which the first seven residues of mature ClpP are removed) [19] was a generous gift from Prof. Walid Houry and Dr. Anna Gribun (University of Toronto). The plasmid was transformed into SG1146GaBL21(DE3) cells (a clpp$^-$ strain) and purified in the same way as ClpP-His$_6$ (see Appendix B).

Synchrotron hydroxyl radical footprinting

Buffer conditions were optimized to minimize quenching of hydroxyl radicals while maintaining ClpA and ClpP solubility and activity. The optimal buffer conditions were found to be 50 mM sodium cacodylate, pH 7.0, 400 mM KCl, and 20 mM MgCl$_2$ (rxn buffer). ClpP maintains full proteolytic activity in this buffer system (as compared to standard Tris or HEPES buffer conditions), while ClpA maintains 70-80% of its ATPase activity. ClpA and ClpP were purified as described above, except after purification and before storage, buffer exchange into rxn buffer was performed using a PD10 column (GE Healthcare) according to the manufacturer’s instructions. The enzymes were then aliquoted, flash-frozen in liquid nitrogen and shipped overnight in dry ice to Brookhaven National Laboratory (Upton, NY), where they were immediately stored at -80°C. For radiolysis, the proteins were diluted to a concentration of 2 mM, then incubated in 1 mM ATPγS for five min prior to exposure to 0-200 ms of mirror-focused [48] polychromatic x-rays at the X28C beamline of the National Synchrotron Light Source using a KinTek quench-flow apparatus (according to published procedures.
Methionine amide was added to a final concentration of 10 mM to prevent secondary oxidation of methionine residues [50]. Exposed samples were digested overnight at 37°C using sequencing-grade modified trypsin (Promega, Madison, WI) and analyzed via LCMS and MSMS as previously described [33, 51] to determine the extent and location of modification. The rate constant for reduction of the fraction of unmodified peptide with exposure was determined via a chi-squared fit to a single exponential function.

**Preparation of ssrA-dabsyl conjugate**

The ssrA peptide (AANDENYALAA) was synthesized by the MIT Biopolymers Facility, and a dabsyl group was added to the N-terminus using dabsyl-chloride (Anaspec). The reaction was performed in 66% acetone and 33% bicarbonate buffer (200 mM sodium bicarbonate, pH 9.0). A 3-fold molar excess of dabsyl-chloride was used, and the reaction was allowed to proceed for 30 min at 80°C. The conjugate was purified by HPLC using a reverse-phase C-phenyl semi-preparatory column (SB-Phenyl, 5 μm, 9.4 x 250 mm, Agilent). The peptide conjugate was stored dry at -20°C and solubilized in DMSO immediately prior to use.

**Peptide degradation assays**

All reactions were performed at 37°C in buffer containing 50 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM MgCl₂, 1 mM DTT, and 10% glycerol. All assays were performed in triplicate, except ssrA-dabsyl experiments, which were performed in duplicate. For insulin chain B degradation, 1 μM tetradeacmer was used for reactions with ClpPΔN or wild-type ClpP alone; for reactions with ClpAP, 0.1 μM ClpP₁₄ and 0.2 μM ClpA₆ were used, and 2 mM ATPγS was added to the reaction mixture. For both SLY-AMC (N-
Succinyl-Leu-Tyr-7-amido-4-methylcoumarin) and ssrA-dabsyl degradation, 0.125 μM ClpP$_{14}$ or ClpPΔN$_{14}$ was used; for degradation by ClpAP, 0.25 μM ClpA$_6$ and 2mM ATPγS were added to the reaction. Hydroxylamine (Aldrich) was added to the reaction mixture for trapping experiments. For insulin chain B and ssrA-dabsyl degradation, 7-amino-4-methyl-coumarin (AMC, Sigma) was added as an internal standard (for HPLC analysis). For reactions with SLY-AMC, degradation was monitored as an increase in AMC fluorescence (ex: 345 nm, em: 440 nm) using a microplate spectrofluorimeter (Molecular Devices, Spectramax Gemini XS). Relative fluorescence units were converted into AMC concentration using a standard curve; addition of hydroxylamine did not affect the AMC standard curve (data not shown). For insulin chain B and ssrA-dabsyl degradation, aliquots of the reaction mixture were removed at various time points and quenched by the addition of an equal volume of 7.4 M Gu-HCl. These samples were then analyzed by HPLC using a C18 reverse phase analytical column (Jupiter, 150 x 4.60 mm, 5 μm, Phenomenex). Linear fits of product amplitude vs. time were used to determine degradation rates except where distinct rapid and slow phases were observed. In that case, plots were fitted to the sum of an exponential and a linear function to obtain both an initial rate for the rapid phase and a steady-state rate for the slow phase.

**MALDI (Matrix Assisted Laser Desorption/Ionization) mass spectrometry**

MALDI spectra were recorded on a PerSeptive Biosystems Voyager-DE STR in positive ion mode using α-cyano-4-hydroxycinnamic acid as the matrix. The instrument was calibrated before each use.

**II.F. Acknowledgments**
We thank Dr. Janna Kiselar for assistance with MS and Dr. Sayan Gupta for footprinting advice; Mike Sullivan, John Toomey and Don Abel for expert technical support for the X28C beamline; Prof. Walid Houry and Dr. Anna Gribun for the plasmid encoding ClpPAN; Dr. Anna Gribun for ClpPAN purification advice; and Profs. Bob Sauer and Tania Baker for the ClpA expression plasmid.

II.G. References


Nature 403, 800-805.


Chapter III: ClpP hydrolyzes a protein substrate processively and independently of the ClpA ATPase: Mechanistic studies of ATP-independent processive proteolysis

*The deconvolution work described in this chapter was performed by Dr. Desmond Lun and Prof. Muriel Médard of the MIT Laboratory for Information and Decision Systems.

III.A. Summary

ATP-dependent proteases are processive, meaning that they degrade full-length proteins into small peptide products without releasing large intermediates along the reaction pathway. In the case of the bacterial ATP-dependent protease ClpAP, ATP hydrolysis by the ClpA component has been proposed to be required for processive proteolysis of full-length protein substrates. We present here data showing that in the absence of the ATPase subunit ClpA, the protease subunit ClpP can degrade full-length protein substrates processively, albeit at a greatly reduced rate. Moreover, the size distribution of peptide products from a ClpP-catalyzed digest is remarkably similar to the size distribution of products from a ClpAP-catalyzed digest. The ClpAP- and ClpP-generated peptide product size distributions are well fitted by a sum of multiple underlying Gaussian peaks with means at integral multiples of ~900 Da (7-8 amino acids). Our results are consistent with a mechanism in which ClpP controls product sizes by alternating between translocation in steps of 7-8 (± 2-3) amino acids and proteolysis. On the structural/molecular level, the step size may be controlled by the spacing between the ClpP active sites, and processivity may be achieved by coupling peptide bond hydrolysis to the binding and release of substrate and products in the protease chamber.

Note: A manuscript describing this work entitled “Jennings, L.D., Lun, D.S., Medard, M., & Licht, S. ClpP hydrolyzes a protein substrate processively and independently of the ClpA ATPase: Mechanistic studies of ATP-independent processive proteolysis” has been accepted for publication in Biochemistry.
III.B. Introduction

Energy-dependent proteases are large molecular machines responsible for the intracellular degradation of misfolded and short-lived regulatory proteins. Members of this family include the eukaryotic 26S proteasome and the bacterial enzymes ClpAP, ClpXP, HslVU, Lon, and FtsH. In these systems, AAA+ (ATPases associated with a variety of cellular activities) components bind, unfold, and translocate protein substrates into barrel-shaped compartmental proteases where the substrates are hydrolyzed and released as peptide products [1-4]. While much attention has been devoted to understanding how protein substrates are recognized, unfolded, and translocated by the ATPase component of energy-dependent proteases [5, 6], much remains to be learned about how proteolysis proceeds after the substrates reach the proteolytic chamber. In particular, the way that the ATPase and protease components may work together to maintain processivity is incompletely understood [7].

One way to study the proteolytic mechanism is to examine the size distribution of peptide products generated by these proteases. Different proteolytic mechanisms make distinct predictions about the shape of the product size distribution and its sensitivity to perturbations in enzymatic function [8]. Size exclusion chromatography allows the product size distribution to be estimated [8-10], with the caveat that the chromatogram reflects both the true size distribution and chromatographic factors that broaden the observed peaks [11]. Previous work has shown that the size range of products generated by energy-dependent proteases spans a 10-fold range (from ~3 residues to ~30 residues). Interestingly, all ATP-dependent proteases for which product size has been characterized [including the 20S archaeal [9] and eukaryotic proteasomes [10, 12], the 26S proteasome
(which includes the protease subunits in the 20S proteasome, accessory ATPase subunits, and other regulatory subunits), ClpAP [8], HslVU [13], and Lon [14, 15]) share this size range of products. Furthermore, both ClpAP and the proteasome have been shown to generate approximately log-normal peptide product size distributions with a peak between 6-9 amino acids and a tail skewed towards longer products [8-10].

The similarity in the range of product sizes and in the shape of product size distributions suggests that all ATP-dependent proteases share a common proteolytic mechanism. However, this mechanism remains a matter of debate. The distance between protease active sites corresponds to 7-8 amino acids [16-18], matching the peak observed at this product length in both proteasome- and ClpAP-generated size distributions. This observation led to the hypothesis that the coordinated action of adjacent active sites controls peptide product sizes (i.e., the “molecular ruler” mechanism) [16, 19]. However, a substantial fraction of both proteasomal [9, 10] and ClpAP [8] products are significantly longer or shorter than 7-8 amino acids, contrary to the prediction of the “molecular ruler” mechanism. Furthermore, both proteasome- and ClpAP-generated peptide product size distributions are unaltered by partial inactivation of up to 70% of the active sites [8, 10], in contrast to the “molecular ruler” mechanism, which predicts that partial inactivation will increase the average product size.

Alternatives to the “molecular ruler” mechanism have been put forward. For both the proteasome and ClpP, the action of product exit pores as a filter has been proposed to account for product size control. This “filter” mechanism postulates that once a substrate enters the chamber, proteolysis continues randomly until the peptide products are small enough to diffuse out of the chamber [10, 18]. For ClpAP, we have proposed a
mechanism in which alternation between translocation and proteolysis controls product sizes. In this mechanism, product sizes are independent of both the proteolysis rate and the translocation rate, consistent with experimental observations [8].

All energy-dependent proteases characterized to date are processive, meaning that they completely degrade protein substrates into small peptides without the release of large intermediates [15, 20-22]. This feature is necessary for product size control in all of the proteolytic mechanisms that have been proposed. Furthermore, the specific mechanism of processive proteolysis is expected to influence the distribution of product sizes produced, but it is not yet clear how processivity is achieved in these systems. Surprisingly, the ATPase component is not always necessary for processivity. For example, the 20S proteasome can degrade full-length proteins into small peptide products (of the same size distribution as that generated by the ATPase-incorporating 26S proteasome [10]) without the aid of an ATPase component [9, 10, 12]. However, for bacterial energy-dependent proteases, it has generally been proposed that the ATPase component is necessary for processive degradation of full-length proteins. While ClpP has been shown to degrade large polypeptides (up to 30 amino acids) processively, the enzyme has not been shown to degrade (processively or otherwise) full-length proteins in the absence of a partner ATPase [20]. The Lon protease has been shown to degrade a full-length protein to a low extent in the absence of ATP, but the only peptides detected were from the terminal regions of the substrate protein, suggesting that this ATP-independent proteolysis is not processive [14].

The mechanism for ATP-independent processive proteolysis is not yet clear. One possibility is that peptide bond hydrolysis provides the driving force for processivity.
This reaction is exergonic [23] and could drive processive proteolysis either through a power stroke or a ratchet mechanism. Peptide bond hydrolysis has been shown previously to drive processive movement of activated collagenase (MMP-1) on collagen fibrils. Processivity in this system seems to be maintained via a Brownian ratchet mechanism, with biased diffusion of the collagenase enzyme on collagen fibrils fueled by proteolysis [24]. A similar proteolysis-driven mechanism might account for the ATP-independent processivity observed with the 20S proteasome.

Given the apparent mechanistic similarities between the proteasome and ClpAP, it is surprising that processive proteolysis of large protein substrates by ClpAP appears to have an absolute requirement for ATP hydrolysis, while a similar reaction can be catalyzed by the 20S proteasome without concomitant ATP hydrolysis. However, since previous work did not put a lower limit on the rate of ClpP-catalyzed degradation of large protein substrates [20], one possibility is that ClpP alone does catalyze processive proteolysis, but at a rate lower than the limit of detection of previous experiments. In that case, the mechanistic similarities between the proteasome and ClpAP would extend to ATP-independent processive cleavage of protein substrates.

In this work, we investigate two questions related to ClpAP-catalyzed processive proteolysis. The first is whether ClpA is required for processive proteolysis of casein, a natively unfolded substrate. The results indicate that, contrary to what might have been predicted based on previous work, the protease subunit ClpP can degrade casein processively in the absence of ClpA, albeit at a greatly reduced rate (a ~2000-fold rate decrease). This observation leads to the second question, which is how the product size distribution of ClpP products compares to the size distribution of ClpAP products. We
show that the size distribution of peptide products from a ClpP-catalyzed digest is remarkably similar to the size distribution of products from a ClpAP-catalyzed digest. These results indicate that ClpP by itself can carry out processive proteolysis and control the product size distribution. Quantitative analysis of the product size distributions provides further mechanistic details. A deconvolution algorithm [11] is applied to the size-exclusion chromatograms in order to extract the true size distributions from underlying chromatographic factors. These deconvolved size distributions quantitatively fulfill the predictions of a mechanism in which ClpP translocates protein substrates in steps of 7-8 amino acids (+/- 2-3 amino acids) and partitions after each step between another translocation step and proteolysis. The physical basis for this stepping mechanism is not yet clear, but we propose a model in which the step size is controlled by the spacing between the active sites, and binding and release of the substrate and products in the protease chamber are coupled to peptide bond hydrolysis.

III.C. Results

Degradation by ClpP is processive

In order to understand the mechanism of product size control by ClpAP, it is important to know whether or not processive proteolysis by this enzyme is ClpA-dependent. Previous studies have shown that ClpP alone can processively degrade large (up to 30 amino acid long) peptide substrates [20], but to our knowledge, ClpP-catalyzed degradation of a full-length protein in the absence of ClpA or ClpX has not previously
been reported. We therefore assessed the ability of ClpP alone to carry out processive degradation of a full-length protein substrate.

In order to facilitate digestion of a full-length protein in the absence of an ATPase, we used as a substrate the natively unfolded protein α-casein (methylated at the lysine side chains to prevent reaction with fluorescamine), and we allowed the degradation reaction to proceed for 24 hours. We then assessed degradation by quantifying the formation of peptide products via their fluorescamine reactivity [25]. The degradation rate was ~2000-fold slower for ClpP alone than for ClpAP; a control digestion using the catalytically inactive S97A ClpP mutant [26] showed no measurable activity for this mutant, indicating that the proteolysis observed with wild-type ClpP was due to the purified enzyme rather than a contaminating protease (Figure III.1.A). In order to assess processivity we measured the ratio of total product formation to casein degradation over time. If degradation is processive, the ratio \( n \) of peptide products formed to intact casein degraded should remain constant over time, due to the fact that formation of \( n \) equivalents of new peptide products will require degradation of one equivalent of casein. However, if degradation is non-processive, the ratio of peptide products formed vs. casein molecules degraded will increase over time; partially degraded intermediates released at early extent of reaction will become substrates for the enzyme, leading to the formation of new peptide products without the degradation of any additional intact casein [21].

The results show that the ratio of peptide products to casein molecules degraded remains constant during ClpP-catalyzed degradation, indicating that degradation by ClpP is processive (Figure III.1.A). Furthermore, since the size distribution of ClpP-generated
ClpP alone can processively degrade a protein substrate. Shown are the (A) ClpP- and (B) chymotrypsin-catalyzed degradation kinetics of casein. Peptide product formation (blue) is shown on the left y-axis, and loss of intact casein (red) is shown on the right y-axis. To show that degradation is ClpP dependent, we measured peptide product formation with the ClpP S97A active site mutant (blue triangles, part A). Error bars represent the standard error of the mean of three independent trials.
products shows a dominant peak around 7-8 amino acids, processive degradation of a ~250 amino acid long substrate such as casein is expected to yield ~30 peptides per casein molecule degraded, even at low extent of reaction. Our results indicate that the ratio of peptide products formed to casein molecules degraded is ~30 throughout the reaction, consistent with processivity. In contrast, in control studies with the non-processive enzyme chymotrypsin, the ratio of peptide products to casein molecules increases over time from 2 at ~10% extent of reaction to 5 at ~40% extent of reaction, as expected for a non-processive enzyme (Figure III.1.B). These data thus support the surprising conclusion that although in the absence of ClpA, ClpP-catalyzed degradation of a full-length unfolded protein substrate is slow, ClpA is not needed to achieve processive proteolysis of this substrate.

*Size-exclusion chromatography of ClpAP products*

In order to better understand the role of ClpA in determining the size distribution of ClpAP-generated peptide products, we analyzed ClpAP’s peptide products using size-exclusion chromatography. The size-exclusion chromatogram of ClpAP-generated peptide products was previously characterized using post-column fluorescamine derivatization to quantify the peptide products [8]. This technique produced a chromatogram with a resolution high enough to resolve the dominant features (such as a peak at 6-8 amino acids and the skewness of the chromatogram toward larger peptide sizes). However, the necessity of pooling fractions for a post-column assay limited the resolution of the data, leaving open the possibility that the size distribution had additional unresolved features. In order to obtain a higher resolution chromatogram, we derivatized ClpAP-generated peptide products with fluorescamine prior to size-exclusion
chromatography [12] and detected the products using an on-line fluorimeter. We used as substrates α-casein and GFP-ssrA. Despite the distinct properties of these two substrates (in fold and amino acid composition), the high resolution size-exclusion chromatograms of both sets of peptide products are remarkably similar. As expected, both chromatograms show a peak at ~800-900 Da (~7-8 amino acids). However, in both cases the higher-resolution chromatogram reveals a previously undetected shoulder peak around 1600-1800 Da (~14-16 amino acids) and a tail skewed towards longer products (Figure III.2 and Figure III.3).

Size-exclusion chromatograms of ClpP products

To help determine how processive proteolysis occurs without substrate translocation by ClpA, we characterized the size distributions of ClpP-generated peptide products in the absence of ClpA. As in the processivity studies, we used casein as a substrate and allowed the reaction to proceed for 24 hours. The size distribution of ClpP-generated peptide products is qualitatively different from the distribution observed for ClpAP products. It shares the same dominant peak around 800-900 Da and a shoulder peak around 1700 Da, but has a significantly shorter tail (products > 2500 Da constitute ~3% of the ClpP-derived size distribution while products > 2500 Da constitute ~10% of the ClpAP products; see Figure III.2).

Validation of MALDI MS as a technique for measuring number distributions of peptide product sizes

The size-exclusion data indicate that the size distribution of ClpP-generated peptide products has a similar overall shape but a narrower range than that derived from
Size-exclusion chromatograms of casein peptide products. A) Size-exclusion chromatograms of ClpAP- and ClpP-derived peptide products from casein digests. Shown are the average traces of three independent trials. B) Calibration curve used to convert retention time to molecular weight. Standard peptides were chromatographed in duplicate or triplicate; error bars represent standard deviation. C) Size-exclusion chromatograms in which the x-axis has been converted from retention time to molecular weight using the calibration curve shown in B.
Size-exclusion chromatogram of ClpAP-derived peptide products from a GFP-ssrA digest. Shown is the average trace of three independent trials.
ClpAP. In order to test this conclusion by an independent method, we used MALDI mass spectrometry to determine the range of peptide product sizes. We digested casein with ClpAP or ClpP, and obtained the MALDI mass spectra of the resulting peptide products. Because ionization efficiencies differ for different peptides, ion intensities in MALDI mass spectra cannot be used to compare amounts of individual peptides. However, counting the number of distinct peaks in a defined size window gives the number distribution of peptide products, which is useful for determining the mass range of peptide products from a given digest.

To assess the utility of MALDI mass spectra in determining number distributions of peptide product sizes, we obtained MALDI-derived number distributions of peptide products from both tryptic and chymotryptic digests. These experimental distributions were compared to the expected number distributions, which were calculated using the well-defined residue specificities of the proteases. In order to obtain a sufficient number of distinct peptides for statistical analysis, we digested bovine serum albumin (BSA) and ClpA together as a mixture, and analyzed the peptide products by MALDI. A histogram plotting the number of peptides detected versus the molecular weight was compared to the expected histograms for both a chymotryptic digest and a tryptic digest of the two proteins (Figure III.4). A chi-squared analysis revealed that the expected chymotrypsin distribution is a good fit to the observed chymotrypsin distribution (0.95>p>0.90), and the expected trypsin distribution fits the observed trypsin distribution (p=0.9). However, the expected chymotrypsin distribution does not fit the observed trypsin distribution (p<0.005), nor does the expected trypsin distribution fit the observed chymotrypsin distribution (p<0.025). Moreover, the number of peptides detected using MALDI was
MALDI analysis accurately reproduces a known number distribution. Shown are the expected peptide number distributions for both a chymotryptic (A) and a tryptic (C) digest of bovine serum albumin (BSA) and ClpA (generated using ExPASy PeptideCutter tool, www.expasy.org/tools/peptidecutter/). The observed peptide distributions found using MALDI for a chymotryptic (B) and a tryptic (D) digest of BSA and ClpA are also shown. Error bars represent the standard deviation of three trials. Because singly-charged ions dominate MALDI spectra [46], M/Z ≈ MW (Da). When the expected and observed distributions for a chymotryptic digest of BSA and ClpA were compared using a chi-squared goodness of fit test, the observed distribution was found to fit the expected distribution (0.95>p>0.90). Similarly, the observed tryptic distribution fit the expected tryptic distribution (p<0.9). Contrarily, a chi-squared analysis found that the observed chymotryptic distribution did not fit the expected trypsin distribution (p<0.025), and the observed tryptic distribution did not fit the expected chymotryptic distribution (p<0.005).
approximately the same as the number of peptides expected (for trypsin, 81 are expected and 81±1 are found; for chymotrypsin, 61 are expected and 70±4 are found). These results indicate that MALDI mass spectrometry allows accurate measurement of the product number distribution for two proteases with distinct specificities, and suggests that this technique will be generally useful in determining number distributions. Use of this technique in combination with liquid chromatography/tandem mass spectrometry might be useful in measuring both substrate specificity (via product identification) and product number distributions.

**MALDI number distributions of ClpAP and ClpP digests**

MALDI was used to determine the number distribution of peptide products generated in ClpAP and ClpP digests of casein. The distributions follow the same pattern observed with size-exclusion chromatography. The distribution obtained from ClpAP digests of casein contains a peak at 1100-1300 Da and a second peak at 2100-2300 Da (Figure III.5.A). However, the distribution obtained from a ClpP digest of casein only shows a peak at 1100-1300 Da; the peak at 2100-2300 Da observed in the ClpAP distribution is absent from the ClpP-derived distribution (Figure III.5.B). These data provide additional evidence showing that the size distribution of ClpP-generated products has a similar overall shape but a narrower range than the size distributions of ClpAP products.

*Fitting multiple Gaussian peaks to the size-exclusion chromatograms*

Because ClpP-catalyzed casein degradation is also processive in the absence of ClpA, comparing the ClpAP size distribution with that of ClpP alone may reveal
MALDI-derived number distributions. Shown are the (A) ClpAP- and (B) ClpP-generated peptide product number distributions from casein digests. Error bars represent the standard error of the mean of three independent trials. Because singly-charged ions dominate MALDI spectra [46], $M/Z \approx MW$ (Da).
mechanistic similarities or differences between ClpA-assisted and ClpA-independent processive proteolysis. In order to compare the size distributions, we fitted the size-exclusion chromatograms to sums of Gaussian components (Figure III.6). The first and largest amplitude component is at ~900 Da. The second component, both broader and smaller in amplitude than the first, is at ~1750 Da and is the major contributor to the shoulder in the chromatogram. A third component, even broader and smaller than the first two, is needed to account for the tail observed in the ClpAP distribution (Figure III.6.A). In contrast, this third component is not needed to account for the ClpP size distribution (Figure III.6.B).

**Deconvolution of the size-exclusion chromatogram**

The size-exclusion chromatograms are determined by both the true size distribution and by chromatographic features that are independent of the size distribution, such as peak broadening and tailing. In order to derive mechanistic conclusions from the size-exclusion data, it would be useful to deconvolve the true size distribution from these chromatographic factors. We therefore developed a deconvolution algorithm for the size-exclusion chromatograms. First, we recorded chromatograms for twelve standard peptides with molecular weights varying from ~400 to ~2500 Da. These chromatograms provided an average peak shape function, the mean peak width of a single peptide, and an estimate for the uncertainty in peak width. Using these values, we carried out a maximum likelihood estimation of the peptide size distributions underlying the observed chromatograms; the method is described in full detail in a separate work [11]. A standard statistical criterion for model discrimination, the Akaike information criterion [27], was used to determine the most probable number of components underlying the
Fits of the size-exclusion chromatograms and the deconvolved size distributions to a sum of Gaussians. The size-exclusion chromatograms of (A, C) ClpAP-, (B, D) ClpP-derived peptide products (from casein digests) were fitted to sums of underlying Gaussian components. The first component is shown in blue, the second in red, and the third in green. In C and D, the raw chromatograms are shown as colored traces, the deconvolved size distributions are shown as black squares, and the fits of the deconvolved distributions to sums of Gaussians are shown as black lines. The underlying Gaussian components are shown as black dotted lines.
chromatograms. This criterion provides an unbiased estimate of the information divergence between the true probability distribution of a system and the estimated distribution derived from a model.

The deconvolved size distributions are qualitatively similar in breadth and skewness to the original chromatograms (Figure III.6.C and III.6.D). This observation shows that the dominant contributor to the breadth of the chromatograms is the true size distribution rather than chromatographic factors. It also indicates that the skewness towards longer products observed in the raw chromatograms is not an artifact of chromatography but rather a feature of the actual product size distributions.

Fitting the deconvolved size distributions

The deconvolved size distributions for ClpAP and ClpP peptide products were fitted to sums of Gaussians. As found for the direct fits to the size-exclusion chromatograms, three Gaussian components were needed to fit the ClpAP deconvolved size distribution, but only two components were necessary to fit the ClpP size distribution (Figure III.6.C and III.6.D). Three parameters were determined for each Gaussian component: the mean, the standard deviation, and the amplitude (Table III.1). In the context of the proposed mechanism for processive proteolysis (Figure III.7), these three parameters can be used to determine the translocation step size, the variability in step size, and the probability that the substrate is cleaved rather than taking a subsequent translocation step (Figure III.8). The determination of the model parameters from the fits is described in the Discussion.
### Table III.1. Results from fitting ClpAP and ClpP deconvolved size distributions to a sum of Gaussian components.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$ (Da)</th>
<th>$\sigma$ (Da)</th>
<th>Relative Amplitude</th>
<th>Average peptide size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClpAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 1</td>
<td>931</td>
<td>324</td>
<td>58%</td>
<td>1408</td>
</tr>
<tr>
<td>Component 2</td>
<td>1801</td>
<td>545</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>Component 3</td>
<td>2724</td>
<td>905</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>ClpP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 1</td>
<td>899</td>
<td>277</td>
<td>72%</td>
<td>1136</td>
</tr>
<tr>
<td>Component 2</td>
<td>1745</td>
<td>599</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>Component 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
A proposed stepwise mechanism for translocation and proteolysis by ClpP. (A) Following binding of the unfolded protein substrate, ClpP translocates this substrate one step (equal to 7-8 amino acids). Following this step, the enzyme partitions between taking another step and proteolysis. Proteolysis after one step results in a pool of peptides at the step size of the enzyme, while proteolysis after two steps results in a pool of peptides at twice the step size, and proteolysis after three steps generates peptides at three times the step size. The partitioning ratio between stepping and proteolysis determines the relative amplitudes of the product pools. For example, a partitioning ratio of 50/50 (as shown) results in product pools with relative amplitudes of 50%, 25%, 12.5%, etc. (B) The mechanism in part A results in a size distribution composed of three Gaussian peaks.
Fits of the deconvolved size distribution data to the predictions of the stepwise model. Blue = ClpAP, Red = ClpP. A) The slope of peak position vs. peak number gives a best-fit value for the step size of 900 ± 20 Da (7.6 ± 0.2 amino acids) and 845 Da (7.1 amino acids) for ClpAP and ClpP, respectively. B) The slope of standard deviation vs. peak number gives a best-fit step size standard deviation of 290 ± 40 Da (2.4 ± 0.3 amino acids) and 320 Da (2.7 amino acids) for ClpAP and ClpP, respectively. C) Fitting the relative amplitudes of the underlying peaks to a geometric distribution \( f(n)=(1-p)^n p \), where \( p \) = the probability of proteolysis and \( n \) = peak number) gives a best-fit value for the probability of proteolysis of 0.57 ± 0.03 and 0.71 ± 0.06 for ClpAP and ClpP, respectively.
III.D. Discussion

ClpP controls processivity and peptide product sizes

We present here data showing that ClpP in the absence of ClpA can degrade the full length (~250-amino acid) natively unfolded protein, casein, albeit at a much slower rate than ClpAP (~2000-fold slower). Despite the low rate of reaction, degradation by ClpP is processive, like ClpAP-catalyzed degradation. These surprising results indicate that ATP-driven translocation by ClpA is not necessary to achieve processive proteolysis of a full-length protein, and suggest that ClpP alone can drive processive substrate translocation, albeit at a very slow rate.

Deconvolved size distributions of ClpAP- and ClpP-derived peptide products provide information about the roles of ClpA and ClpP in product size control. Each size distribution has a large peak at 800-900 Da (7-8 amino acids) and a second smaller peak at 1700-1800 Da (14-15 amino acids). In addition, the ClpAP distribution has a tail (comprising ~10% of the distribution) extending out to ~4000 Da; this tail is largely missing from the ClpP distribution. MALDI-derived number distributions confirm that the longest products present in the ClpAP distribution are not present in the ClpP distribution. Like the processivity data, these results indicate that, contrary to our previous hypothesis [8], the action of ClpA is not required to control product sizes. Instead, it appears that product sizes are controlled largely by ClpP.

ClpAP and ClpP size distributions are consistent with a stepping mechanism

Both the raw size-exclusion chromatograms and the deconvolved size distributions of ClpAP- and ClpP-derived peptide products can be fitted well with a sum
of underlying Gaussian peaks. A size distribution composed of one or more Gaussian peaks is consistent with a mechanism in which ClpP-catalyzed deterministic translocation of the substrate into the protease chamber alternates with proteolysis at one or more of the ClpP active sites. In such a mechanism, if a translocation event is always followed by a proteolysis event, one Gaussian peak would be predicted in the size distribution. Furthermore, the mean of this Gaussian peak would reveal the average length of substrate threaded into the active sites during each translocation event (termed the “step size” of the enzyme), while the standard deviation would reveal the variation in this step size. A size distribution composed of multiple peaks (as observed) could arise if the enzyme partitions between translocation and proteolysis, rather than cleaving the substrate after each translocation event. Therefore, a second peak in the size distribution would arise from partitioning twice toward translocation before proteolysis, and a third peak would arise from partitioning three times toward translocation prior to proteolysis, etc (Figure III.7).

*Obtaining best-fit values for step size, standard deviation in step size, and partitioning ratios*

Best-fit values for three parameters of the stepwise mechanistic model outlined in Figure III.7 (the step size, the standard deviation in the step size, and the probability of translocation versus proteolysis) were obtained from fits of the ClpAP- and ClpP-derived deconvolved size distributions to sums of Gaussian components. These three parameters were determined from three predictions that the model makes about the properties of the product size distribution. First, the model predicts that the step size can be determined from the means (peak positions) of the underlying Gaussians. Specifically, the model
predicts that the underlying components will be regularly spaced, and that the spacing between the components will equal the translocation step size of the enzyme. For ClpAP- and ClpP-derived peptide product size distributions, the underlying Gaussian components are spaced ~800-900 Da (7-8 amino acids) apart. The best-fit slope of the peak position versus peak number gives a step size of 7-8 amino acids for both ClpAP and ClpP (Figure III.8.A).

In addition, the stepwise model (Figure III.7) predicts that the standard deviation in the translocation step size can be determined from the standard deviation of the underlying Gaussian components. In this model, the standard deviations of the underlying Gaussians will increase linearly with each component. When some intrinsic variation in size is associated with each translocation step, taking two translocation steps before proteolysis will lead to a doubling of the associated variability, taking three translocation steps before proteolysis will triple the variability, etc. As predicted by the model, the observed standard deviations of the underlying Gaussian components of both the ClpAP- and ClpP-derived distributions increase linearly with peak position, and the slope of the standard deviation versus peak number gives a best-fit value for the standard deviation in the step size of 2-3 amino acids for both ClpAP and ClpP (Figure III.8.B).

Finally, the stepwise mechanism outlined in Figure III.7 predicts that the partitioning ratio between proteolysis and further translocation can be determined from the amplitudes of the underlying Gaussian components. Specifically, the proposed mechanism predicts that the amplitudes of the underlying Gaussian components will decrease according to a geometric distribution [28], \( f(n) = (1-p)^{n-1}p \), where \( p \) = the probability of proteolysis and \( n \) = peak number. This relationship is observed in the data,
and fitting the underlying peak amplitudes to a geometric distribution gives a best-fit value for the probability of proteolysis vs. translocation for both enzyme conditions. The best-fit probability of proteolysis is 0.57 and 0.71 for ClpAP and ClpP respectively (Figure III.8.C). This slight increase in partitioning towards proteolysis for the ClpP-catalyzed reaction versus the ClpAP-catalyzed reaction is consistent with the deficiency of longer ClpP products as observed by both size-exclusion chromatography and MALDI mass spectrometry. The small differences between these distributions are somewhat difficult to interpret mechanistically. The similarities, however, are striking. Although the proteolytic rate is slowed tremendously by the absence of ClpA, the lack of an ATPase does not prevent processive proteolysis in discrete steps.

Alternative mechanisms for product size control

The size distribution data are inconsistent with two alternative mechanisms for peptide product formation. Product size might be determined by a "filter" mechanism, in which larger products cannot diffuse out of the complex and are therefore re-cleaved until they are small enough to escape [10, 18]. However, it is difficult to reconcile this mechanism with the observation that the peaks in the product size distribution are approximately Gaussian. A structural pore that acts as a filter would be expected to admit smaller peptides preferentially, skewing the size distribution toward smaller products. In addition, since exit through equatorial pores appears to be the dominant mode of peptide release for ClpP [29, 30], the filter mechanism seems most consistent with only a single peak in the product size distribution (corresponding to the equatorial pore size), in contrast to the observation of up to three peaks experimentally. Another alternative mechanism is the "molecular ruler," in which cleavage at two adjacent active
sites produces products at a size equivalent to the inter-site distance [19]. However, in order to achieve high fractions of product at the size corresponding to the inter-site distance, this mechanism requires that all the active sites cleave with high efficiency. This mechanism thus seems to be inconsistent with the formation of the largest products, which would require a relatively high probability that one-third or one-half of active sites failed to cleave their substrates. The requirement for cleavage at two adjacent active sites is also inconsistent with previous work indicating that the peptide size distribution is not altered by partial inactivation of the ClpP active sites [8].

_A physical model for a ClpP-controlled stepwise translocation mechanism_

While the kinetic and product size distribution results presented here support a mechanism in which substrate translocation occurs in discrete steps, the specific physical basis for these steps remains open to debate. The similarity in product size distributions for ClpAP- and ClpP-catalyzed caseinolysis suggests that, like the ClpAP-catalyzed reaction [31], the ClpP-catalyzed reaction proceeds via processive translocation of the polypeptide substrate. Although direct proof that the polypeptide enters the ClpP pore in discrete steps remains to be established, this proposal provides an economical explanation for the processivity and size-distribution data presented here. While it is not yet clear how stepwise translocation might be coordinated in the absence of an ATPase, the crystal structure of ClpP may provide a useful starting point for thinking about possible mechanisms, since it shows the arrangement of the active sites and substrate binding sites. First, the spacing between the active sites may help determine the size of the steps. The crystal structure of ClpP indicates a spacing of about 7-8 amino acids between the active sites [17, 18], consistent with the spacing between the Gaussian components. In addition,
the active sites of each heptameric ring are connected via a substrate-binding groove containing the S and S' pocket sites [32]. Each active site is expected to act as a relatively high affinity binding site due to the presence of these substrate binding regions [17, 18, 32]. Taking into account these structural features, a physical model for processive translocation can be proposed that accounts for processive proteolysis and product size control by ClpP via switches in binding affinity that are coupled to active site occupancy and peptide bond hydrolysis (Figure III.9).

In the proposed model (Figure III.9), tight binding (i.e., complete binding to both the S and S' pocket sites) is necessary for efficient proteolysis (Figure III.9, step 2). Therefore, once a free segment of peptide binds tightly to both the S and S' pockets of an active site, peptide bond hydrolysis occurs (Figure III.9, step 3), generating a small (~8 amino acid) product at the S' site and leaving the larger product at the S site. Following hydrolysis, the affinity of both products is assumed to decrease, allowing them to leave their respective sites. Rapid release of the smaller product leaves the S' site open (Figure III.9, step 4). The equilibrium for productive translocation of the remaining protein substrate is favorable, since translocation causes both the S and S' sites to be occupied. A translocation step can thus occur (Figure III.9, step 5). Translocation is assumed to occur through essentially one-dimensional diffusion through the substrate binding groove, since that mode of transport would maximize favorable enzyme-substrate contacts during translocation. The translocation step can then be followed by tight binding of the translocated substrate to the S and S' sites (Figure III.9, step 6), leading to another round of proteolysis. Alternatively, if stabilization at the S' site fails to occur rapidly enough compared to release of the translocated segment from that site (Figure
Speculative model for ClpP-controlled stepping mechanism. Depicted in blue is a ring of active sites from one ClpP heptameric unit. The active sites are represented as stars and are connected via the substrate binding groove. A) A close-up view of the active site binding region boxed in B. The S1-S8 and S1′-S8′ binding sites of one active site (depicted in red) are shown. B) After the substrate binds to ClpP, it translocates via diffusion into the active site region (step 1). Once the substrate binds completely to an active site (i.e., to all the S and S′ binding sites) the interaction between substrate and enzyme is stabilized (step 2), leading to a high affinity binding state from which proteolysis occurs (step 3). Proteolysis produces an ~8 amino acid long peptide product in the S′ site. (In order to illustrate this mechanism we assume that translocation proceeds from the C-terminus to the N-terminus [39], placing the smaller product at the S′ site.) After proteolysis and product release (step 4), the substrate-enzyme binding affinity is lowered, and a translocation step can once again occur (step 5). This translocation step can be followed by stabilization (step 6) to the high affinity binding state (from which proteolysis occurs), or the substrate can dissociate (step 7) before stabilization. Dissociation can be followed by an additional translocation step (step 8). After this step, the enzyme again partitions between stabilization (step 9), and ultimately proteolysis, and dissociation (step 10). Proteolysis after this second translocation step (step 8) will lead to a peptide product ~16 amino acids in length.
III.9, step 7), a second translocation step can take place before the next peptide hydrolysis reaction (Figure III.9, step 8). If stabilization at the S' site (Figure III.9, step 9) again fails to occur rapidly enough compared to release of the translocated segment from that site (Figure III.9, step 10), a third translocation step can take place before proteolysis, etc. This mechanism is similar to the "Brownian ratchet" mechanisms that have previously been proposed for classical motors [33].

The proposed model accounts for processive proteolysis and the generation of peptide products with average lengths at integral multiples of 7-8 amino acids. In the proposed model, the irreversible peptide bond cleavage and product release steps ensure that translocation is processive and unidirectional by driving the catalytic cycle in a single direction. The step size of 7-8 amino acids is determined by the size of the peptide binding sites in the protease active site [32], with the assumption that once both the S and S' peptide binding sites are completely occupied, the substrate binds tightly, allowing proteolysis to occur. To the extent that the switch to tight binding is incompletely coupled to complete occupancy of the peptide binding sites, the step size will be variable, as observed experimentally. When part of the substrate fails to bind tightly before a second translocation event can occur, the result is an increase in product length by 7-8 residues; each additional translocation event extends the product length by an additional increment of 7-8 residues. Further experimental work will be required to test this model. However, it has the advantage of accounting for existing results without requiring amide bond hydrolysis to drive large conformational changes in ClpP, a possibility that structural studies suggest is unlikely [34].
The mechanism described above is also consistent with previous reports indicating that the size distribution of peptide products is unaltered by inactivation of 70% of the active sites in ClpP [8]. Unlike the molecular ruler mechanism, this mechanism only requires one functional active site in the chamber to achieve processivity. Assuming that cleavage products can dissociate rapidly and exit the chamber, rather than re-binding to another active site, the mechanism in Figure III.9 predicts that partial inactivation will not change the size distribution of products, since the translocating substrate only interacts with a single active site. Of course, the presence of multiple active sites is expected to accelerate the rate of the initial productive encounter between the substrate and an active site, consistent with the observation that covalent modification of 70-90% of the active sites does decrease the steady-state reaction rate [8, 20].

In the ClpAP complex, the conformational changes that drive translocation through the ClpP central pore may be coupled to the conformational changes that drive ClpA-catalyzed substrate translocation through the ClpA central pore [35]. ClpA by itself can translocate substrates at rates much faster than those achieved by ClpP. However, the similarity between the product size distributions observed for ClpAP and ClpP alone suggests that ClpA does not change the elementary events that mediate substrate translocation through ClpP. Coordination of ClpA and ClpP conformational changes would allow ClpA to accelerate translocation through ClpP without changing the basic ClpP mechanism. Ample experimental evidence exists for conformational coupling between the protease and ATPase components of Clp proteases [36-38]. In particular, the ClpP N-terminus may act to transduce conformational changes in ClpA to ClpP. The ClpP N-terminal domain is necessary for interaction with ClpA [30, 39], and the binding
of ClpA causes the ClpP N-terminus to leave the central pore of the complex (see Chapter 2 of this thesis). The presence or absence of the N-terminus in the ClpP central pore also appears to have a profound influence on the reactivity of the active site serine (see Chapter 2 of this thesis), even though these regions of the complex are separated by tens of angstroms [17]. The catalytic cycle proposed for ClpP-mediated translocation may thus be tightly coupled to the rapid conformational changes involved in ClpA-mediated translocation [31, 35, 40], allowing fast processive proteolysis only when both components are present.

These results may also provide insights into the mechanisms of other ATP-dependent proteases. The eukaryotic proteasome has been shown to processively degrade protein substrates in the absence of its ATPase binding partner [21], and the size distribution of its peptide products exhibits multiple peaks [12]. The current work provides analytical tools for evaluating mechanisms for proteasome-catalyzed processive proteolysis. Deconvolution of size-exclusion chromatograms to provide size distributions could, in principle, allow additional testing of the proposed “filter” mechanism [10] and alternative mechanisms. Further investigation of these mechanisms may help provide new insights into the important question of how epitope peptides are generated in the immune system.

III.E. Materials and Methods

Protein purification
ClpA, ClpP-His$_6$, and GFP-ssrA were purified as described previously [8, 41-43] (see also Appendices A-C). ClpPS97A-His$_6$ was purified from SG1146GaBL21(DE3) cells (a clpp- strain) in the same way as ClpP-His$_6$. ClpP-His$_6$ was purified from either DH5$\alpha$ cells or S1146GaBL21(DE3) cells; the two preparations had comparable activity (see Figure III.10). $\alpha$-casein and chymotrypsin were purchased from Sigma.

**Figure III.10**

**Comparison of casein degradation activity of ClpP purified from two different strains of E. coli.** Red: ClpP purified from DH5$\alpha$ cells. Green: ClpP purified from SG1146GaBL21(DE3) cells (a clpp' strain). The peptide products formed from the degradation of casein were quantified using fluorescamine. Error bars represent the standard deviation of three independent trials.
Reductive methylation of substrate proteins

To prevent reaction of fluorescamine with lysine side chains, reductive methylation of GFP-ssrA and α-casein was carried out using formaldehyde and sodium cyanoborohydride as previously described [8, 44]. All degradation assays were performed using methylated substrates. However, for simplicity, methylated GFP-ssrA and methylated α-casein are referred to here as “GFP-ssrA” and “casein”, respectively.

Processivity measurements

Casein was degraded by ClpP or ClpPS97A in buffer containing 50 mM HEPES, pH 7.5, 20 mM MgCl₂, 100 mM NaCl, 10% glycerol, 0.5 mM DTT, 100 μM casein, and 4 μM wild-type ClpP₁₄ or ClpPS97A₁₄. The reactions were allowed to proceed for 24 h at 37°C. Chymotrypsin-catalyzed degradation of casein was performed under the same conditions, except 0.133 μg/mL chymotrypsin was used in place of ClpP, and the reaction was allowed to proceed for 2 h. Aliquots of the reaction mixtures were taken at various time points and quenched by the addition of an equal volume of 7.4 M Gu-HCl. Peptide product formation was measured using fluorescamine [25], which forms a fluorescent product upon reaction with primary amines. Each quenched sample (10 μL) was mixed with 100 μL of 100 mM HEPES, pH 6.8, and 50 μL of 0.3 mg/mL fluorescamine in acetone (freshly prepared before use). After 7.5 min, the fluorescence was measured (ex: 370 nm, em: 510 nm) using a microplate spectrofluorimeter (Molecular Devices Spectramax Gemini XS). To convert relative fluorescence units into peptide product concentration, a standard curve was prepared using the ssrA peptide (AANDENYALAA) derivatized with fluorescamine in the same way as the digested
peptides. To measure casein degradation, each quenched sample was loaded onto an HPLC reverse-phase C18 analytical column (Jupiter, 150 x 4.60 mm, 5 μm, Phenomenex). ClpP digests were eluted with the following gradient program at a flow rate of 1 mL/min: isocratic elution with 95% solvent A (0.1% TFA in water)/5% solvent B (0.1% TFA in acetonitrile), 5 min; linear gradient from 95% A/5% B to 80% A/20% B, 5 min; linear gradient from 80% A/20% B to 30% A/70%B, 30 min. For chymotrypsin digests, the following gradient program was used: isocratic elution with 95% A/5% B, 5 min; linear gradient from 95% A/5% B to 80% A/20% B, 15 min; linear gradient from 80% A/20% B to 50% A/50% B, 60 min. Casein eluted as a single peak that was well separated from product and protease peaks, and was detected by absorbance at 280 nm. Experiments were performed in triplicate.

**Digestion of casein and GFP-ssrA by ClpAP or ClpP for size-exclusion chromatography studies**

Proteolysis of GFP-ssrA by ClpAP was performed in a reaction mixture containing 50 mM HEPES, pH 7.5, 30 mM MgCl₂, 300 mM NaCl, 10% glycerol, 0.5 mM DTT, 25 μM GFP-ssrA, 50 nM ClpP₁₄, 100 nM ClpA₆, 10 mM ATP, and an ATP regenerating system consisting of 30 mM phosphocreatine and 0.05 units/μL creatine phosphokinase (from rabbit muscle, Sigma). The total reaction volume was 300 μL. The reaction was allowed to proceed for 2 h at 37°C and was quenched by boiling. For proteolysis of casein by ClpAP, the same reaction conditions were used with the exception that 20 mM MgCl₂ and 100 mM NaCl were present in the buffer, and ClpP₁₄ (100 nM) and ClpA₆ (200 nM) were used to degrade casein (25 μM). The reaction was
allowed to proceed for 45 min at 37°C, at which point the reaction was quenched by boiling. Casein was degraded by ClpP in a reaction mixture containing 50 mM HEPES, pH 7.5, 20 mM MgCl₂, 100 mM NaCl, 10% glycerol, 0.5 mM DTT, 50 μM casein, and 5.5 μM wild-type ClpP_{14}. These reactions were allowed to proceed for 24 h at 37°C, followed by quenching via boiling. In each case, controls were performed in the absence of enzyme, and the experiments were performed in triplicate.

Size-exclusion chromatography of peptide products

Peptide products were derivatized with fluorescamine prior to chromatography to enable on-line detection. Following desalting with a reverse-phase cartridge (Sep-Pak C18, Waters) and concentration by centrifugal evaporation, peptide products were resuspended in 400 μL of 100 mM HEPES, pH 6.8. 3-20 μL of peptide products (diluted to a total volume of 20 μL with 100 mM HEPES, pH 6.8) were then added to 10 μL of 0.3 mg/mL fluorescamine (in acetone, freshly prepared before use). After 1 min, 30 μL of water was added to the reaction, and the sample was loaded immediately onto the size-exclusion column [12]. Size-exclusion HPLC was performed using a polyhydroxyethyl aspartamide column (200 mm × 4.6 mm, PolyLC) [45]. The mobile phase was 200 mM sodium sulfate, 5 mM potassium phosphate, 25% acetonitrile (pH 3.0, adjusted with phosphoric acid), and the flow rate was 0.125 mL/min. Peptides were detected using an on-line fluorescence detector (ex: 380 nm, em: 510 nm). To determine the apparent molecular weight of the peptides eluted, the column was calibrated with 5-12 commercially-available standard peptides in the 400-2500 Da range that had been derivatized with fluorescamine in the same way as the digest peptides.
Chymotrypsin digest of BSA and ClpA

Two nmol of protease-free BSA (bovine serum albumin, inhibited with Protease Inhibitor Cocktail, Sigma) along with 2 nmol ClpA were denatured by boiling in the presence of 8 M urea for 15-20 min. The mixture was then diluted until the urea concentration was 2 M, and ammonium bicarbonate was added to a final concentration of 100 mM. 7.5 µg sequencing grade, modified chymotrypsin (Princeton Separations) was then added, and the digestion was allowed to proceed for 6 h at 30°C. After this time both proteins had been degraded to >95% completion as judged by reverse-phase chromatography.

Trypsin digest of BSA and ClpA

Two nmol of protease-free BSA (inhibited with Protease Inhibitor Cocktail, Sigma) along with 2 nmol ClpA were denatured by boiling in the presence of 3 mM DTT and 6 M guanidine hydrochloride for 15-20 min. The mixture was then diluted until the guanidine hydrochloride concentration was 0.85 M, and ammonium bicarbonate was added to a final concentration of 50 mM. Twelve µg sequencing grade, modified trypsin (Promega) was then added, and the digestion was allowed to proceed for 17 h at 37°C. After this time both proteins had been degraded to >97% completion as judged by reverse-phase chromatography.

Matrix Assisted Laser Desorption Ionization (MALDI) mass spectrometry

Peptide products were desalted using a reverse-phase cartridge (Sep-Pak C18, Waters) and concentrated by centrifugal evaporation if necessary. One µL (~100 pmol) was spotted onto a MALDI sample plate and allowed to dry; thereafter 0.6 µL of matrix (15 mg/mL α-Cyano-4-hydroxycinnamic acid in 80% acetonitrile) was spotted on top of
the peptides. The MALDI spectra were recorded on a PerSeptive Biosystems Voyager-DE STR in positive ion mode. The instrument was calibrated each time before use. In order to count the number of peptides in each spectrum, the background was first corrected using the MATLAB (MathWorks, Natick, MA) command msbackadj. A noise cutoff was set (at $\mu + 2.5\sigma$ ion counts, where $\mu =$ mean ion count of the background signal, and $\sigma =$ standard deviation of the background signal), and all peaks above this threshold were counted as peptide products.

*Deconvolution of size-exclusion chromatograms*

The average peak shape and standard deviation in peak width for single peptides were determined from measurements of 12 standard peptides. A random-restart gradient algorithm was used to generate a maximum likelihood estimate for the size distribution given the observed chromatogram and the properties of peaks arising from single peptides [11]. The computation was implemented in MatLab 7.1 (scripts available on request).

**III.F. Acknowledgments**

We thank Dr. Bob Sauer, Dr. Tania Baker, and Dr. Søren Molin for donating plasmids; Mary Lee of the Sauer lab for purification advice; Dr. Barbara Imperiali for generously providing the use of the HPLC; Dr. Pete Wishnok and Dr. Steve Tannenbaum for use of their mass spectrometry facility.
III.G. References


Chapter IV: Synchrotron protein footprinting supports substrate translocation by ClpA via ATP-induced movements of the D2 loop

*This work was performed in collaboration with Dr. Jen Bohon and Prof. Mark Chance of Case Western Reserve University and Christine Phillips of the MIT Department of Chemistry.

IV.A. Summary

Synchrotron x-ray protein footprinting is used to study structural changes upon formation of the ClpA hexamer. Comparative solvent accessibilities between ClpA monomer and ClpA hexamer samples are in agreement throughout most of the sequence with calculations based on two previously proposed hexameric models. The data differ substantially from the proposed models in two parts of the structure: the D1 sensor 1 domain and the D2 loop region. The results suggest that these two regions can access alternate conformations in which their solvent protection is greater than in the structural models based on crystallographic data. In combination with previously reported structural data, the footprinting data provide support for a revised model in which the D2 loop contacts the D1 sensor 1 domain in the ATP-bound form of the complex. These data provide the first direct experimental support for the nucleotide-dependent D2 loop conformational change previously proposed to mediate substrate translocation.

IV.B. Introduction

ATP-dependent proteases are responsible for a variety of essential cellular regulatory functions, the most notable of which are the dissolution of protein aggregates and the degradation of unwanted proteins; both of these processes are required for cell growth, mediation of stress responses, and protein quality control. Typically, the unfolding and degradation of protein substrates proceeds via ATP-dependent translocation of these substrates from the ATPase component through a narrow pore into the degradation chamber of the protease component. Understanding the conformational changes that enable ATP-dependent proteases to manipulate their substrates is a useful first step in designing agents to modulate bacterial physiology by activating [1] or inhibiting these molecular machines; however, the relevant conformational changes are not yet well-defined. In this study, we investigate the arrangement of protein domains in the pre-hydrolytic state of the energy-dependent protease ClpA.

ClpAP is a roughly cylindrical *E. coli* ATP-dependent protease complex comprised of a double-ringed tetradecameric ClpP protease core flanked by a hexameric ClpA ring on one or both ends [2]. The ClpA chaperone is required for substrate recognition, unfolding and translocation of the unfolded substrate to the ClpP active sites. Natural substrates for ClpAP include the plasmid P1 replication initiator RepA, the heme biosynthetic enzyme HemA and a number of carbon starvation proteins [3-5]. In addition, ClpAP is able to degrade proteins with one of two types of identifying ‘tags.’ The first type of marker is the identity of the N-terminal amino acid. N-terminal arginine, lysine, leucine, phenylalanine, tyrosine and tryptophan all target a protein for N-end rule degradation [6]. The second type of tag is the 11-amino acid ssrA sequence
(AANDENYALAA), the addition of which can target stably folded proteins to ClpAP for ATP-dependent degradation, including tagged lambda repressor and tagged GFP [7, 8].

ClpA undergoes self-assembly from the monomeric into the active hexameric form in the presence of ATP or the non-hydrolyzable analog ATPγS. Hexamerization is required for both substrate binding to ClpA and for the formation of the ClpAP complex [9, 10]; the ATPγS-bound form of the ClpA hexamer is competent to activate ClpP for the degradation of large peptide substrates (~30 residues), although ATP hydrolysis is required for efficient proteolysis of protein substrates [11]. ATP hydrolysis is required for ClpA-catalyzed unfolding of large proteins and for translocation of all substrates except small peptides. Protein substrates typically require several rounds of ATP hydrolysis for complete conversion into peptide products; the ClpAP complex remains associated throughout this process and through several rounds of degradation [12].

A 3-D reconstruction at a resolution of 29 Å via cryo-electron microscopy [13] indicates that the macromolecular ClpAP complex forms three compartments: the digestion chamber inside ClpP, a small compartment between ClpA and ClpP, and a chamber inside of ClpA. More detailed structural information is available for the individual components of the structure in the form of crystal structures of the ClpP tetradecamer [14-16] and of the ClpA monomer [17]. In addition, based on the ClpA monomeric crystal structure, two ClpA hexameric models have been published (Figure IV.1.A and IV.1.B). The first model, published by Guo et al. [17], was constructed using crystal structures of the hexameric forms of NSF-D2 and HslU as templates. The second model, published by Hinnerwisch et al. [18], was constructed using the p97 hexameric crystal structure as a template. With the exception of the ClpA N-domain, which is
Figure IV.1.

A

ClpP Binding Side

B

ClpP Binding Side

C

ADP-bound

ATP-bound
Hexameric models and proposed D2 loop movement. D1 sensor and D2 loop are highlighted in magenta and green, respectively. A) A cut-away side view of Hexamer 1, the ClpA hexameric model published by Maurizi and coworkers [17], on the left, with a zoom of the pore region on the right. The D1 sensor of chain F and the D2 loop of chain A are highlighted, with probe residues Y324, P537, P538, Y540, and F543 shown in blue. B) A cut-away side view of Hexamer 2, the ClpA hexameric model published by Horwich and coworkers [18], on the left, with a zoom of the pore region on the right. The D1 sensor and D2 loop from chain A and the same probe residues as in A are highlighted. C) Proposed movement of the D2 loop upon ATP hydrolysis with the D1 sensor and D2 loop colored as in A and B. The D2 loop is in the “down” conformation in the ADP-bound post-hydrolytic state. The D2 loop moves to the “up” conformation upon rebinding of ATP (the pre-hydrolytic state).
missing from the second model, the two models are qualitatively very similar to one another; each includes a central pore and is consistent with the structure inferred from electron microscopy. However, little direct, high-resolution structural information is available concerning the ClpA hexamer structure, nor have these models been tested against experimental data.

The details of the hexameric model are mechanistically important because large conformational changes in ClpA have been proposed to mediate unfolding of protein substrates and their translocation through the ClpAP complex [18]. Recent results with photoreactive substrates indicate that ClpA’s D2 loop contacts the substrate in the course of translocation. Furthermore, mutations in a region of the protein immediately adjacent to the D2 loop allow substrate binding but not degradation. Based on these results, a mechanism was proposed in which motions of the D2 loop mediate movement of the substrate through the ClpA pore during unfolding and translocation [18]. In the proposed mechanism, the D2 loop binds the substrate in an “up” conformation (Figure IV.1.C). Upon ATP hydrolysis, the D2 loop is proposed to drag the substrate through the central pore of the ClpA hexamer toward the face of the complex that binds ClpP. Once the D2 loop and the substrate are in this “down” conformation, the substrate can be released.

Other functional data are also consistent with the hypothesis that the D2 loop binds substrate tightly in a pre-hydrolytic conformation and releases it in a post-hydrolytic conformation. When ClpA is bound to the poorly-hydrolyzable nucleotide analogue ATPγS, it binds peptide substrates with high affinity [19]. Single-molecule fluorescence experiments indicate that ClpA can assume both high-affinity and low-
affinity conformations, with high-affinity peptide binding favored in the presence of ATPγS and low-affinity binding favored in the presence of hydrolyzable ATP [20].

Existing structural data can be interpreted in terms of a post-hydrolytic “down” conformation, but direct evidence for a pre-hydrolytic “up” conformation has not previously been reported. The hexameric structural models currently in use are derived from a ClpA monomer structure containing bound ADP [17]. These models place the D2 loop and nearby residues close to the ClpP-binding face of ClpA: i.e., in a “down” conformation. These residues are observable in the structure, suggesting that the “down” conformation represents a reasonably stable and well-ordered state of ClpA. Observation of the proposed alternate “up” conformation in the pre-hydrolytic state would provide significant evidence in favor of the proposed mechanism of translocation by D2 loop motions.

To test the hypothesis that the pre-hydrolytic form of the ClpA hexamer places the D2 loop in the “up” conformation, we used synchrotron protein footprinting to investigate the solvent accessibilities of domains in ClpA. Protein footprinting probes the solvent accessibility of side chains in the macromolecule, allowing protein interaction sites and conformational changes occurring upon complex formation to be mapped to specific areas of the protein. Synchrotron protein footprinting works via the generation of large quantities of hydroxyl radicals by direct irradiation of a protein solution over milliseconds. These radicals oxidatively modify protein side chains [21, 22] when they react with solvent-exposed protein regions, while areas of the protein not exposed to solvent are protected from these modifications. Protease digestion of the sample followed by liquid-chromatography-coupled mass spectrometry (LCMS) and tandem
mass spectrometry (MSMS) allows identification of affected regions and often the specific residue that is modified. Changes in modification rate with the addition of complex components identify interprotein interactions and conformational changes.

In this study, we report the results of synchrotron protein footprinting experiments on both disassembled ClpA subunits and the ClpA hexamer bound to the poorly-hydrolyzable nucleotide analogue ATPγS. The footprinting data are most consistent with an ATPγS-bound ClpA structural model in which the D2 loop and associated residues contact the D1 sensor 1 domain, a region within one of ClpA’s Walker A ATPase sites. This conformation would position the D2 loop at the end of the ClpA hexamer closest to the substrate entry site, rather than at the end closest to the ClpP binding site. The footprinting data thus provide new evidence for a pre-hydrolytic “up” conformation of the ClpA D2 loop. These results provide the first structural support for a previously proposed mechanism in which translocation by ClpA is mediated by movements of the D2 loop.

IV.C. Results

Synchrotron x-ray footprinting was carried out under two sets of conditions: in the absence of nucleotide, where ClpA is disassembled, and in the presence of ATPγS, where it is hexameric [9, 10]. MSMS-verifiable coverage of the ClpA tryptic digest encompassed 78% of the protein (Figure IV.2). For those peptides not observed in the mass spectrum, no information can be concluded. Ten tryptic peptides were identified in the ClpA monomer samples with modification confirmed by MSMS; seven of these also displayed modification in hexamer samples. Relative modification rates are listed in
<table>
<thead>
<tr>
<th>Peptide Segment</th>
<th>Color Coding</th>
<th>Amino Acid</th>
<th>Spectrum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLNQELELSL NMAFARAREH RHEFMTVEHL LLALLSNPSA REALEACSVD</td>
<td>Black: MSMS identified peptides</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>LVALRQELEA FIEQTTFVLP ASEEEERDTQ TLSFQRVLQR AVFHVQSSGR</td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>NEVTGANVLV AIFSEQESQA AYLLRKHEVS RLDVVFISH GTRKDEPTQS</td>
<td>Blue: peptides with modification</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>SDPGSQPNSE EQAGGEERTE NFTTNLNLQ A RVGGGDPLIG REKELEAIQ</td>
<td>Red: specifically identified modified residues, Grey: not identifiable in spectra. Red boxes indicate areas of nucleotide interaction, green boxes indicate residues implicated in substrate binding and translocation, and the black box indicates the ‘IGL’ sequence required for binding to ClpP.</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>VLCRRKNALDLVGSNGVGK TAIAEGLAVR IQVGDPEVM ADCTIYSLDI</td>
<td></td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>GSSLAGTRF GDFEKRFKAL LKQLEQDTNS ILFIDEIHTI IGGAASGGQ</td>
<td></td>
<td></td>
<td>300</td>
</tr>
<tr>
<td>VDAANLIKPL LSSGKIRVIG STTYQEFNSI FEKDRALARR FQKIDITEPS</td>
<td></td>
<td></td>
<td>350</td>
</tr>
<tr>
<td>IEETVQIING LKPYEEAHHD VRYTAKAVRA AVELAVKYN DRHDLDKAI</td>
<td></td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>VIDEAGARAR LMPVSRRKRT VNVADIESVV ARATIKEKS VSQSDRTLK</td>
<td></td>
<td></td>
<td>450</td>
</tr>
<tr>
<td>NLGDRKMLV FQOEKAIEAL TEAIKMARAG LGHEHKPVGS FLFAGPTGVG</td>
<td></td>
<td></td>
<td>500</td>
</tr>
<tr>
<td>KTEVTVQLSK ALGIELRFQD MSEEYERTHTV SPLICAPPQG VGFQGGLLT</td>
<td></td>
<td></td>
<td>550</td>
</tr>
<tr>
<td>DAVIKHPHAV LLLDEIEKAH PDVFNLLLQV MDNGTLDNN GRKADFERNV</td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>LVMTTNAGVR ETERKD G HQDNSSTAME EIKIFTPPEF RNRDLNNWF</td>
<td></td>
<td></td>
<td>650</td>
</tr>
<tr>
<td>DHLSTDVREQ VVDKFIVELQ VQLDQKVSYL EVSQEARNWL AEKGYDRA</td>
<td></td>
<td></td>
<td>700</td>
</tr>
<tr>
<td>ARPMAVQD NLKPLANEL LFGLSVDGGQ VTVALDEK F ELTYGFQSAQ</td>
<td></td>
<td></td>
<td>750</td>
</tr>
<tr>
<td>KHKAEEAH</td>
<td></td>
<td></td>
<td>758</td>
</tr>
</tbody>
</table>

**Tryptic peptide coverage of ClpA.** Color coding of amino acids: Black: MSMS identified peptides, Blue: peptides with modification, Red: specifically identified modified residues, Grey: not identifiable in spectra. Red boxes indicate areas of nucleotide interaction, green boxes indicate residues implicated in substrate binding and translocation, and the black box indicates the ‘IGL’ sequence required for binding to ClpP.
Table IV.1, and the locations of the peptides in the monomer crystal structure and in the hexamer models are shown in Figure IV.3. Although the same concentration of protein (2 μM) was used for both ClpA monomer and hexamer experiments, 1 mM ATPγS was required for the formation of the hexamers. This ATP analogue provides a significant quenching effect (factor of 4.35) on the dose received by the hexamer samples due to scavenging of the hydroxyl radicals; all hexamer rates stated are therefore normalized for this effect (Figure IV.4).

The peptides analyzed in this experiment cover a variety of regions of the protein thought to be important for function. In each region, the experimental protection factors can be compared with predictions based on solvent accessibility analysis of the published monomeric structure or the two reported hexameric models (Tables IV.1 and IV.2). For most residues, the agreement is reasonably good, but significant deviations are observed in a number of areas (Figure IV.3). Two of the anomalous regions, the D2 loop (corresponding to peptide 533-555) and the D1 sensor 1 (corresponding to peptide 318-333), are particularly noteworthy in terms of the magnitude of the effects observed and the possible mechanistic implications; however, it is important to discuss the entire set of footprinting observations in order to place the D2 loop and D1 sensor 1 results in the proper context.

The peptide comprised of amino acids 458-465 contains the VFGQD sequence considered to be responsible for nucleotide binding in the D2 AAA+ domain [17]. While peptide side chain solvation calculations using either of the previously published hexameric models predict that this region of the protein becomes more accessible upon hexamer formation, the footprinting experiments indicate that it is slightly less accessible
Table IV.1. Rate constants for the modification of monomeric and hexameric ClpA

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>Oxidation Rate Monomer</th>
<th>Oxidation Rate Hexamer</th>
<th>Monomer/Hexamer Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>77-86</td>
<td>DTQPTLSFQR</td>
<td>0.3±0.03</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>144-168</td>
<td>KDEPTQSSDPGSQPNSEEQAGGEER</td>
<td>1.7±0.10 (1.3±0.09)</td>
<td>(1.3)</td>
<td></td>
</tr>
<tr>
<td>169-181</td>
<td>TENFTTNLQNLAR</td>
<td>0.43±0.096</td>
<td>-</td>
<td>(1.3)</td>
</tr>
<tr>
<td>221-230</td>
<td>TAIAEGLAGWR</td>
<td>0.6±0.04 (0.8±0.65)</td>
<td>(0.7)</td>
<td></td>
</tr>
<tr>
<td>318-333</td>
<td>VIGSTTYQESFNIFEK</td>
<td>1.7±0.42 (0.2±0.17)</td>
<td>(9.8)</td>
<td></td>
</tr>
<tr>
<td>458-465</td>
<td>MLVFGQDK</td>
<td>16.9±0.51 (8.2±0.82)</td>
<td>(2.0)</td>
<td></td>
</tr>
<tr>
<td>519-527</td>
<td>FDMSYMER</td>
<td>113.8±14.93 (0)</td>
<td>(0)</td>
<td></td>
</tr>
<tr>
<td>533-555</td>
<td>LIGAPPGYVGFQDDQGGLTDAVIK</td>
<td>6.0±0.72 (0.5±0.27)</td>
<td>(10.9)</td>
<td></td>
</tr>
<tr>
<td>616-633</td>
<td>SIGLIHQDNSDAMEEIK</td>
<td>30.5±2.29 (9.6±0.37)</td>
<td>(3.2)</td>
<td></td>
</tr>
<tr>
<td>740-751</td>
<td>NELTYGFQSAQK</td>
<td>0.6±0.11</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

MS/MS verified oxidation sites are in bold and underlined.
Numbers in parentheses are normalized for ATPyS quenching (factor of 4.35).
Table IV.2. Calculated peptide solvent accessibility values for structural models of ClpA

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Domain</th>
<th>Monomer&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MonoΔN&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Hexamer1&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Hexamer2&lt;sup&gt;d&lt;/sup&gt;</th>
<th>FP&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Monomer/Hexamer1&lt;sup&gt;a&lt;/sup&gt; Ratio</th>
<th>Monomer/Hexamer2&lt;sup&gt;b&lt;/sup&gt; Ratio</th>
<th>Monomer/FP Model&lt;sup&gt;c&lt;/sup&gt; Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>77-86</td>
<td>N</td>
<td>651.3</td>
<td>-</td>
<td>713.4</td>
<td>-</td>
<td>-</td>
<td>0.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AA168</td>
<td>N-D1 loop</td>
<td>194.7</td>
<td>194.7</td>
<td>73.6</td>
<td>194.3</td>
<td>194.4</td>
<td>2.65</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>169-181</td>
<td>D1</td>
<td>890.0</td>
<td>890.0</td>
<td>744.2</td>
<td>789.9</td>
<td>789.8</td>
<td>1.20</td>
<td>1.13</td>
<td>1.13</td>
</tr>
<tr>
<td>221-230</td>
<td>D1 near WA</td>
<td>271.0</td>
<td>297.9</td>
<td>271.7</td>
<td>298.5</td>
<td>298.5</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>318-333</td>
<td>D1 sensor</td>
<td>711.2</td>
<td>711.2</td>
<td>676.2</td>
<td>521.2</td>
<td>307.3</td>
<td>1.05</td>
<td>1.36</td>
<td>2.31</td>
</tr>
<tr>
<td>458-465</td>
<td>D2 nucl. int.</td>
<td>470.5</td>
<td>470.5</td>
<td>511.6</td>
<td>488.7</td>
<td>488.8</td>
<td>0.92</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>519-527</td>
<td>D2</td>
<td>527.7</td>
<td>527.7</td>
<td>218.4</td>
<td>401.4</td>
<td>484.7</td>
<td>2.42</td>
<td>1.31</td>
<td>1.09</td>
</tr>
<tr>
<td>533-555</td>
<td>D2 loop</td>
<td>1318.3</td>
<td>1318.3</td>
<td>855.0</td>
<td>702.7</td>
<td>591.9</td>
<td>1.54</td>
<td>1.88</td>
<td>2.23</td>
</tr>
<tr>
<td>616-633</td>
<td>D2</td>
<td>828.8</td>
<td>828.8</td>
<td>773.5</td>
<td>867.8</td>
<td>869.3</td>
<td>1.07</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>740-751</td>
<td>D2 C-term</td>
<td>852.2</td>
<td>852.2</td>
<td>902.4</td>
<td>849.5</td>
<td>849.7</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>KSF [17]; <sup>b</sup>KSF without N-domain; <sup>c</sup>Guo et al. model [17]; <sup>d</sup>Hinnerwisch et al. model [18]; <sup>e</sup>footprinting model.

Italics indicate missing or disparate residues; Amino acid 168 is Ala in Hexamer1 and Arg in the monomer, the Hexamer2 model and the FP model. Aside from amino acid 168, peptide 144-168 is absent from the models. Residues 616-623 are also missing from the models; the accessibilities are calculated using only residues 624-633.
Comparison of experimental ClpA monomer/hexamer modification rate ratios and expected solvent accessibilities. Grey indicates a lack of coverage via MSMS. Blue indicates peptides for which only the unmodified peptide was identified via MSMS. Data/model comparison color coding: Green, relative agreement; Yellow, modest differences; Red, major disagreement. Specific modified residues are shown as spheres. A) ClpA monomer crystal structure 1KSF [17]. B) Hexamer 1 model [17]. C) ClpA monomer crystal structure 1KSF with the N-domain removed for comparison to D), Hexamer 2 model [18].
Figure IV.4.

Dose-response curves for ClpA peptide modification as a function of x-ray exposure time. ClpA monomer curves are shown in black, and ClpA hexamer curves are shown in blue. Grey lines indicate hexamer values normalized for the presence of ATPγS.
in hexameric form. However, because the modified region is thought to participate in nucleotide binding, a process required for hexamer formation, the small increase in protection observed may be attributable to the space taken up by the nucleotide and localized conformational changes associated with the binding of the small molecule.

The IGL triplet is found within peptide 616-629 at positions 617-619. This loop is known to be essential for binding to ClpP [23-25] and is predicted to remain on the surface upon hexamer formation, implying a small decrease in solvent accessibility. The data shows significant protection (>3-fold), indicating a conformational change that buries the probe side chains. It is important to note that residues 616-623, which contain the region of interest, are missing from the ClpA crystal structure (and the models), and thus are not included in the accessibility calculations. Also, the probe residues are H621 and M629, and while these are protected in the hexamer, this may not translate into protection of the IGL triplet, which is expected to remain accessible in order to bind to ClpP.

Peptide 144-168 is almost entirely unobservable in the crystal structure of the ClpA monomer (only residue 168 provided resolvable electron density), indicating that there is a high level of flexibility in this segment. The model places both ends of this fragment at the hexamer surface facing away from ClpP. This peptide shows small but significant protection from modification upon hexamer formation, possibly indicating an increased ordering of the strand as it makes contacts along the surface.

Peptide 77-86 is within the N-domain of the protein, a region thought to be fairly flexible. The rate of modification of this peptide is quite low, even in the monomer. Although no modification was evident in hexamer samples, the quenching effect of the
ATPyS may have placed any fragments beyond the detection limits of the technique. Peptide 169-181, found within the D1 domain, exhibits a similar pattern of modification. Peptide 221-230 is proximal to the AAA+ Walker A nucleotide-binding consensus sequence motif in the D1 domain. This region of ClpA shows no difference in modification rate between the monomeric form of ClpA, with no nucleotide bound, and the hexameric form, with nucleotide present. Solvent accessibility calculations are consistent with these observations.

Peptide 519-527, a region immediately N-terminal to the D2 loop, exhibits significant protection upon formation of the hexamer. This region is highly oxidized in the monomer, with a modification rate of >100 s\(^{-1}\); multiple residues are observed to be oxidized in this peptide in the MSMS spectra. Even unexposed monomer samples showed a small amount of oxidation within this peptide. However, hexameric samples showed no oxidation above this background level, even upon greater exposure times, indicating complete burial of the probe residues as a function of complex formation. This is in reasonable agreement with the structural models on the residue level (Table IV.3), which predict that hexamerization will lead to a significant decrease in solvent-accessibility for the M525 probe residue that accounts for the majority of the oxidized product. On the peptide level, the solvent-accessible area in the hexamer is predicted to be relatively small (~300 Å\(^2\)) compared to that of the monomer (~500 Å\(^2\)), but would nevertheless predict a measurable rate of modification for these residues.

The hexameric models’ predictions and the experimental data differ most substantially in two regions of ClpA (Figure IV.3). One is the β-sheet D1 sensor 1
Table IV.3. Calculated probe amino acid solvent accessibility values for structural models of ClpA

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Domain</th>
<th>Solvent Accessibility (Å²)</th>
<th>Monomer</th>
<th>MonoΔN</th>
<th>Hexamer1</th>
<th>Hexamer2</th>
<th>FP</th>
<th>Monomer/ Hexamer1 Ratio</th>
<th>Monomer/ Hexamer2 Ratio</th>
<th>Monomer/ FP Model Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>P80</td>
<td>N</td>
<td>35.7</td>
<td>n/a</td>
<td>44.3</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
<td>0.81</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>F84</td>
<td>N</td>
<td>4.9</td>
<td>n/a</td>
<td>9.3</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
<td>0.53</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>P147, 153,157</td>
<td>N-D1 loop</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>F172</td>
<td>D1</td>
<td>152.6</td>
<td>152.6</td>
<td>40.8</td>
<td>67.1</td>
<td>67.1</td>
<td></td>
<td>3.74</td>
<td>2.27</td>
<td>2.27</td>
</tr>
<tr>
<td>L179</td>
<td>D1</td>
<td>55.4</td>
<td>55.4</td>
<td>36.8</td>
<td>55.9</td>
<td>55.9</td>
<td></td>
<td>1.51</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>W229</td>
<td>D1 near Walker A</td>
<td>93.1</td>
<td>120.0</td>
<td>85.0</td>
<td>120.3</td>
<td>120.3</td>
<td></td>
<td>1.10</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Y324</td>
<td>D1 sensor I</td>
<td>88.0</td>
<td>88</td>
<td>64.8</td>
<td>133.9</td>
<td>26.8</td>
<td>1.36</td>
<td>0.66</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>M458</td>
<td>D2 nucleotide int.</td>
<td>133.7</td>
<td>133.7</td>
<td>136.9</td>
<td>99.0</td>
<td>56.8</td>
<td>0.98</td>
<td>1.35</td>
<td>2.35</td>
<td></td>
</tr>
<tr>
<td>M521</td>
<td>D2</td>
<td>2.1</td>
<td>2.1</td>
<td>23.9</td>
<td>26.8</td>
<td>3.1</td>
<td>0.09</td>
<td>0.08</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Y524</td>
<td>D2</td>
<td>16.1</td>
<td>16.1</td>
<td>18.0</td>
<td>34.6</td>
<td>12.9</td>
<td>0.90</td>
<td>0.46</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>M525</td>
<td>D2</td>
<td>41.3</td>
<td>41.3</td>
<td>16.2</td>
<td>21.2</td>
<td>48.9</td>
<td>2.55</td>
<td>1.95</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>P537</td>
<td>D2 loop</td>
<td>65.9</td>
<td>65.9</td>
<td>103.2</td>
<td>132.6</td>
<td>61.4</td>
<td>0.64</td>
<td>0.50</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>P538</td>
<td>D2 loop</td>
<td>131.1</td>
<td>131.1</td>
<td>32.7</td>
<td>10.5</td>
<td>62.2</td>
<td>4.01</td>
<td>12.53</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td>Y540</td>
<td>D2 loop</td>
<td>26.9</td>
<td>26.9</td>
<td>19.4</td>
<td>14.8</td>
<td>41.8</td>
<td>1.39</td>
<td>1.82</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>F543</td>
<td>D2 loop</td>
<td>97.4</td>
<td>97.4</td>
<td>107.3</td>
<td>4.2</td>
<td>25.4</td>
<td>0.91</td>
<td>23.19</td>
<td>3.83</td>
<td></td>
</tr>
<tr>
<td>H621</td>
<td>D2 near ClpP loop</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>106.9</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>M629</td>
<td>D2</td>
<td>119.0</td>
<td>119.0</td>
<td>138.5</td>
<td>92.6</td>
<td>118.8</td>
<td>0.86</td>
<td>1.28</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Y744</td>
<td>D2 C-term</td>
<td>68.1</td>
<td>68.1</td>
<td>102.6</td>
<td>30.7</td>
<td>68.4</td>
<td>0.66</td>
<td>2.22</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>F746</td>
<td>D2 C-term</td>
<td>67.1</td>
<td>67.1</td>
<td>73.3</td>
<td>61.7</td>
<td>64.9</td>
<td>0.92</td>
<td>1.09</td>
<td>1.03</td>
<td></td>
</tr>
</tbody>
</table>

*aIKSF [17]; bIKSF without N-domain; cGuo et al. model [17]; dHinnerwisch et al. model [18] efootprinting model.

Due to the lack of N-domain in Hexamer2 and the footprinting model, the values are calculated for these models and the respective ratios without this portion of the structure for both monomer and hexamer. Amino acids P147, P153, P157 and H621 are missing from all models as they do not appear in the monomeric crystal structure [17].
sequence (referred to herein as the D1 sensor), contained within peptide 318-333. This peptide becomes highly protected upon hexamer formation (~10-fold). Its modification rate is at the lower limit of detection when ClpA is in the hexameric form, while the rate is easily measurable for the ClpA monomer (Table IV.1). This is in contrast to calculations based on the models, which predict only a very minor decrease in overall solvent accessibility.

The oxidation behavior of peptide 533-555 also deviates from the predicted solvent accessibilities provided by the models. This peptide contains the majority of the D2 loop and all of the amino acid residues within the D2 domain shown to affect substrate binding and/or ClpA function in mutational analyses [18]. The GYVG sequence within this peptide (including one of the oxidized residues, Y540) is highly conserved. It is expected to have particular functional significance based on the observation that the Y540A mutant is able to bind ssrA-linked substrates, but ClpAP proteolytic activity is inhibited [18]. Four different residues are oxidized in this segment during exposure, but the signals are overlapping in the chromatogram, so the individual residue oxidation patterns are not distinguishable. Formation of the ClpA hexamer significantly reduces the solvent accessibility of these amino acids (~10-fold decrease in modification rate). As observed for the D1 sensor, the modification rate for residues 533-555 is close to the lower limit of detection in the hexamer form, in contrast to the substantial modification rate observed for the monomer (Table IV.1). The modest decrease in solvent accessibility predicted for this region is unlikely to account for this level of protection, particularly considering the significant solvent-accessible area (~800Å) predicted to remain in both hexamer models (Table IV.2).
For the majority of the modified peptides, the specific amino acid modified has been identified through tandem mass spectrometry. Despite the general agreement of the footprinting data with the proposed hexameric models when considering overall accessibility for most peptides, the predicted specific amino acid side chain accessibilities are significantly disparate from the data (Table IV.3). As the models are built from monomeric subunits with the goal to satisfy overall domain and general structural requirements observed in similar structures and consistent with cryoEM studies [13, 17], they do not take into account changes at smaller scales that could cause these differences. Using constraints based on the protection maps created by the footprinting data, with the addition of the missing residues and information about their accessibility, the existing ClpA hexameric models can be refined to incorporate these changes.

IV.D. Discussion

Our footprinting results are generally consistent with the previous ClpA structural data and the previously proposed hexameric models. For the peptides analyzed, the regions containing very reactive residues that are predicted to be highly accessible in the hexameric models are generally reactive enough to produce detectable amounts of oxidation products. These regions include the nucleotide binding site and residues at the surface of the hexamer. Those residues for which modification is expected but not observed are predominantly found within the N-domain. This flexible region [26] may assume a different conformation in solution than in the crystal structure. In addition, the residues that change dramatically in accessibility upon hexamer formation mostly lie along the central axis of the hexamer (Figure IV.3). This observation is consistent with
the idea that the intersubunit contacts that define the ClpA central pore help to protect residues in this region from the solvent. Evidence from electron microscopy [2, 13, 27] puts a strong constraint on the global structures that can be considered for ClpA hexamers. The agreement of the footprinting data with the broad outlines of the hexameric model thus supports the utility of this technique in probing the ultrastructure of ClpA.

Previous work does not provide direct evidence for the positioning of individual domains within the hexamer, and it is on this issue that the footprinting data differ from previous models of the hexamer. There are two regions of the protein that react with hydroxyl radical at a moderate-to-high rate in disassembled ClpA subunits and at a low-to-undetectable rate in the ClpA hexamer: the D1 sensor and the D2 loop. These regions are more highly protected than expected based on solvent accessibility calculations of the hexameric models (Figure IV.5). There are two alternative explanations for this observation: either these two regions form a new contact in the ATPγS-bound hexamer, or they insert themselves into other regions of the hexamer that are protected from solvent.

The most parsimonious explanation for the footprinting data is that the D1 sensor domain and the D2 loop contact each other in the ATPγS-bound hexamer. Based on the available crystallographic evidence, this interaction is feasible. A proposed structural model, shown in Figure IV.6.B, was constructed in which the flexible D2 loop contacts the D1 sensor; this model has reasonable bond lengths and bond angles, and does not introduce unfavorable steric interactions. This interpretation of the structural data would be consistent with the proposed mechanism for translocation [18], as it places the
Relative solvent accessibilities in the monomer and hexamer states. For a given set of data, relative solvent accessibilities for the monomeric (m) and hexameric (h) states were parameterized as \([(m/h)-1]/[(m/h)_{\text{max}}-1]\), where \((m/h)_{\text{max}}\) is the maximum monomer/hexamer ratio in the set of data. This parameterization normalizes the largest increase in protection on hexamer formation to a value of 1 and the smallest to a value of 0. For peptide 519-527, where the hexamer accessibility is too low to be measured, the parameter was set to 1. Solid line: experimental modification rates, dashed line: predicted solvent accessibilities from the monomer and hexamer structural models.
Cross section of the ClpA hexamer illustrating the pore region. A) depicts the Hexamer2 model [18] and B) the FP model. The sequences containing the D1 sensor 1 region (318-333) and the D2 loop (526-538) peptides are highlighted and labeled. Coloring is the same as in Figure 3 (Green, relative agreement with footprinting results; Red, major disagreement with footprinting results). In A) the loop is in the “down” position and does not protect residues Y324, Y540, or F543 (blue) from solvent access in the pore. In B) the loop is in the “up” position where it is able to protect Y324, Y540, and F543 from solvent. The FP model (B) provides better agreement with the footprinting results. Bottom figures depict the D1 Sensor 1 and D2 Loop regions in spacefill form.
substrate-binding D2 loop in the proximity of the entrance to the ClpA pore when the complex is in the prehydrolytic state. The revised structural model in Figure IV.6.B still does not account for the very high protection observed for residues 519-527 on the peptide level. However, the specific solvent accessibility of the reactive sulfur of M525 is reduced from 12.3 Å² in the monomeric structure to 1.7 Å² in the revised hexameric model and that of M521 is 0 Å² for both monomer and hexamer structures. At the resolution provided by the current experiments and analysis, it is not possible to rule out a subtle effect of local dynamics as the source of the anomalously high oxidation rate observed for this peptide in the ClpA monomer [28].

An alternative explanation for the footprinting data is that both the D2 loop and the D1 sensor become buried in different solvent-protected regions of the ATPγS-bound hexamer structure. If so, the D2 loop and the D1 sensor would be more solvent-protected than predicted by the original hexameric model, but they would not be in contact with each other. However, it would be difficult to accommodate this possibility without also including large, global changes in the hexameric model. The D2 loop might be able to insert itself into the gap between the D1 and D2 domains on the equator of the complex, but large motions of the D1 sensor would not appear to be possible without a large rearrangement of the structure. Such a rearrangement might be possible, but would be without precedent in existing structural studies of both the hexamer and the monomer. Furthermore, such a rearrangement would likely perturb the solvent accessibility in other regions in addition to the D1 sensor domain; such perturbations are not observed.

When both the footprinting results and previous structural studies are taken into account, the structural model in which the D2 loop contacts the D1 sensor is thus most
consistent with the data. It accounts for large increases in protection for these regions that would be difficult to explain without postulating unprecedented rearrangements of the tertiary structure. Small-scale rearrangements of local structure/chemical environment could account for the effects of hexamerization on residues 519-527 (i.e., the almost complete protection observed for a region predicted to be incompletely buried). It is worth noting that this region is adjacent to the D2 loop. Movement of the D2 loop might cause local structural changes in this region that would account for the protection data, however, the resolution of our data does not allow us to speculate about the nature of such rearrangements. With more sophisticated computational modeling [28, 29], it may be possible to explain the protection observed in this region.

The larger, more easily interpretable effects observed for the D2 loop proper and the D1 sensor provide useful information about state-dependent conformational changes in ClpA. The revised structural model (Figure IV.6.B) is in agreement with a previously proposed mechanism for translocation [18], in which the D2 loop resides near the entry of the central pore before nucleotide hydrolysis (Figure IV.1.C). The previously reported hexameric models might be viewed as a good description of the ADP-bound state of the hexamer. These models are in fact derived from monomer structures that contain bound ADP [17, 18]. With that assumption, the two hexamer models would represent the two key intermediates in the proposed translocation mechanism: the pre-hydrolytic state with the D2 loop “up” and the post-hydrolytic state with the D2 loop “down” (Figure IV.1.C).

In the future, efforts will be made to use transient state and/or single-molecule kinetic experiments to detect the proposed conformational change. Stopped-flow synchrotron footprinting studies have previously been carried out [30, 31], and adapting
them to the ClpA system appears to be feasible. These time-resolved footprinting experiments would allow determination of whether the proposed conformational change occurs and is kinetically competent for translocation. Single-molecule fluorescence experiments have also been carried out on ClpA [20], and might be adapted to allow observation of the proposed conformational change (e.g., by detection of FRET between fluorophores attached to the D2 loop and D1 sensor).

The results of the current study also suggest that synchrotron x-ray footprinting will be generally useful for the study of domain motions in other proteolytic machines and chaperones. Like the Clp proteases, archaeal and mammalian proteasomes must translocate protein substrates through a central pore to present them to protease active sites. The conformational changes involved in translocation by these proteolytic complexes are still incompletely understood; synchrotron footprinting techniques may be useful in addressing the question of which protein domains mediate substrate translocation in these systems and other molecular machines.

IV.E. Materials and Methods

Protein purification

ClpA (with a M169T mutation that enhances solubility and increases levels of full-length protein expression [32]) was purified in the Licht lab as previously described [33, 34] (see Appendix A). After purification and before storage, buffer exchange into RXN buffer (50 mM sodium cacodylate, pH 7.0, 400 mM KCl, and 20 mM MgCl₂) was performed using a PD10 column (GE Healthcare) according to the manufacturer’s instructions. The ATP hydrolysis rate of the enzyme in this buffer system is 70-80% of
the rate observed using standard HEPES buffer conditions (50mM HEPES, pH 7.5, 300mM KCl, 20mM MgCl₂, 10% glycerol, 0.1% nonylphenylpolyethylene glycol). The enzyme stock solution was aliquoted, flash frozen in liquid nitrogen, and shipped overnight in dry ice to Brookhaven National Laboratory (Upton, NY), where it was immediately stored at -80°C.

**Synchrotron hydroxyl radical footprinting**

Samples were thawed on ice and diluted to 2 μM ClpA in RXN buffer. Diluted samples were kept at 4°C prior to and during the experiments and used within 12 h of thawing. All samples other than disassembled ClpA samples (referred to as “monomer,” although dimers and trimers might also be present [32, 35, 36] were incubated in 1 mM ATPγS for 5 min prior to exposure. Exposure conditions were predetermined by following the dose-dependent degradation of the fluorescent compound Alexa® Fluor 488 (Invitrogen, Carlsbad, CA) in the presence of RXN buffer [37]. The ratio of this degradation rate to that of a similar experiment also containing 1 mM ATPγS provided the normalization factor of 4.35 for comparison of monomer and hexamer data. Samples were exposed to a mirror-focused [38] synchrotron x-ray beam (7 mrad angle, focus value of 4.5) at the X28C beamline of the National Synchrotron Light Source at Brookhaven National Laboratory for 0 - 100 ms. The exposure time of the samples was controlled via flow rate through the flow cell in the KinTek (Austin, TX) stopped-flow apparatus [37]. Oxidation was quenched by the addition of methionine amide to a final concentration of 10 mM. Irradiated protein samples were digested with sequencing grade modified trypsin (Promega, Madison, WI) at an enzyme to protein ratio of 1:20 (w/w) at 37°C overnight. The digestion reaction was terminated by adding formic acid to a final
concentration of 0.1%. The resulting peptides (1 pmol) were loaded onto a 300 μm ID x 5 mm C18, PepMap nano Reverse phase (RP) trapping column to pre-concentrate and wash away excess salts using a U 3000 nano HPLC (Dionex, Sunnyvale, CA). The loading flow rate was set to a 25 μL/min, with 0.1% formic acid (pH 2.9) as the loading solvent. Reverse phase separation was performed on a 75 μm ID x 15 cm C18, PepMap nano separation column using nano separation system U 3000 (Dionex). Peptide separation was accomplished using buffer A (100% water and 0.1% formic acid) and buffer B (20% water, 80% acetonitrile and 0.1% formic acid). Proteolytic peptide mixtures eluted from the column with a 2% per minute acetonitrile gradient were introduced into an LTQ FT mass spectrometer (ThermoFisher Scientific, Waltham, MA) equipped with a nano spray ion source and using a needle voltage of 2.2 kV. MS and tandem MS spectra were acquired in the positive ion mode, with the following acquisition cycle: a full scan recorded in the FT analyzer at resolution R=100000 followed by MSMS of the eight most intense peptide ions in the LTQ analyzer. MSMS spectra of the peptide mixtures were searched against an E. coli data base for modifications (oxidation) of the tryptic peptides from the ClpA protein using BioWorks 3.2 software (ThermoFisher Scientific). In addition, detected MSMS mass spectral data for modified peptides were manually interpreted and correlated with hypothetical MSMS spectra predicted for the proteolysis products of the ClpA protein with the aid of the ProteinProspector (UCSF, CA) algorithm. The detected total ion currents were utilized to determine the extent of oxidation by separate quantitation of the unmodified proteolytic peptides and their radiolytic products by dividing the peak area corresponding to the modified peptide by that of the total peptide (modified and unmodified) [39]. Levels of
modification versus exposure time were plotted and fitted with a single exponential [40] via chi-squared minimization to determine the rate constant.

**Solvent accessibility calculations**

Solvent accessible surface area was calculated using the ‘surface’ function in the CCP4 package [41]. The PDB structure 1KSF [17] was used for monomer calculations and the models of the ClpA hexamer used were obtained from the authors of Guo et al. [17] and Hinnerwisch et al. [18].

**Modeling of the D2 loop position**

The proposed loop movement was modeled in COOT [42], and the Regularize function was used to assist in modeling. The Regularize command minimizes the function \( S \), defined as: \( S = S_{bond} + S_{angle} + S_{torsion} + S_{plane} \), thus minimizing the difference between each term and the ideal value for that term [42]. All altered residues (526-540) remained in the Ramachandran-allowed regions.

**IV.F. Acknowledgments**

Center for Synchrotron Biosciences, Center for Proteomics, Case Western Reserve University, is supported by the National Institute for Biomedical Imaging and Bioengineering under P41-EB-01979. The National Synchrotron Light Source at Brookhaven National Laboratory is supported by the Department of Energy under contract DE-AC02-98CH10886. Work in the Licht laboratory is supported in part by a Beckman Young Investigator Award. We thank Drs. Janna Kiselar and Serguei Ilchenko, who performed FTMS for this work; Dr. Wuxian Shi, for assistance with solvent accessibility calculations and Dr. Sayan Gupta for footprinting advice; Mike Sullivan,
John Toomey and Don Abel for expert technical support for the X28C beamline; and Bob Sauer and Tania Baker for the ClpA expression plasmid.

IV.G. References


158
Appendix A: ClpA purification protocol


Expression of ClpA

Overview

The gene encoding ClpA M169T (a mutation eliminating false translational start [1]) was previously cloned into a pET9 vector (containing Kanamycin resistance) and transformed into *E. coli* BL21 cells. In the pET9 plasmid, the clpa gene is under the control of the T7 promoter. *E. coli* BL21 cells contain a chromosomally integrated cassette in which the T7 RNA polymerase is expressed from the *lacUV5* promoter. Upon the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), the T7 RNA polymerase is expressed, inducing expression of ClpA.

Materials

- **LB media with Kanamycin** – 1 L
  - 10 g tryptone
  - 5 g yeast extract
  - 5 g NaCl
  - 1 mL 1M NaOH
  - Fill to 1L with water
  - Autoclave 25 min, 121°C
  - Before starting culture add Kanamycin to 30 μg/mL

Protocol

- Grow 4 × 5 mL cultures in LB media overnight
- Prepare 4 L of LB media for overexpression
- Add 5 mL of overnight starter culture per liter of LB media
- Grow at 37°C to an optical density of ~0.6AU
- Induce expression of ClpA by adding IPTG to a final concentration of 1 mM
- Grow at 25°C for an additional 3 h
- To harvest cells, centrifuge for 15 min at 6000 x g
- Store the pellet at -80°C.
Purification of ClpA

Overview

ClpA is purified from *E. coli* by first lysing the cells using sonication. Nucleic acids are precipitated by the addition of poly(ethyleneimine) (PEI), and ClpA is then precipitated by the addition of more PEI. ClpA is extracted in buffer containing 400 mM KCl, and ammonium sulfate is added to precipitate ClpA.

At this point, ClpA is separated into two populations—those molecules soluble in low (100 mM) salt (termed “low salt” ClpA), and those insoluble at this concentration (termed “high salt” ClpA). “Low salt” ClpA is further purified on a HighS cation exchange column (Bio-rad), and then precipitated by the addition of ammonium sulfate. Following this step, “low salt” and “high salt” ClpA are again separated, and “low salt” ClpA is purified further on a HighQ anion exchange column (Bio-rad). Purified ClpA is then concentrated to > 5 mg/mL. This protocol typically yields 70-75% pure ClpA which can be stably stored at -80°C for several months.

Materials

- **Buffer A0:** 500 mL
  - 50 mM HEPES, pH 7.5
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer A100:** 500 mL
  - 50 mM HEPES, pH 7.5
  - 100 mM KCl
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer A400:** 1 L
  - 50 mM HEPES, pH 7.5
  - 400 mM KCl
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer A1000:** 500 mL
  - 50 mM HEPES, pH 7.5
  - 1000 mM KCl
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer P0:** 500 mL
  - 20 mM potassium phosphate buffer, pH 7.5
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer P100:** 500 mL
  - 20 mM potassium phosphate buffer, pH 7.5
  - 100 mM KCl
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **Buffer P1000:** 500 mL
  - 20 mM potassium phosphate buffer, pH 7.5
  - 1000 mM KCl
  - 2 mM EDTA
  - 10% Glycerol
  - 1 mM DTT (add fresh before use)
- **10% PEI w/v in water titrated to pH 8 with HCl**
- **Solid ammonium sulfate (40% saturation = ~21 g per 100 mL)**
Protocol

*Note: All steps are performed at 4°C unless otherwise noted.

DAY 1

- **Lyse cells**
  - Resuspend cell pellet in 4.3 mL of Buffer AO per g of cells.
  - Sonicate cells 5 x 1 min (2 s on, 1 s off; 70% amplitude).
  - Centrifuge for 30 min at 30,000 x g to clear the lysate; retain supernatant and discard pellet.
  - Save 200 µL of supernatant for **SAMPLE 1** = Crude Extract 1.

- **First PEI Precipitation (to remove nucleic acids)**
  - Add 5 µL of 10% poly(ethyleneimine) (PEI) solution per mL of lysate (creating a final concentration of 0.05% PEI).
  - Mix thoroughly for 20 min at 4°C.
  - Centrifuge the turbid mixture for 30 min at 30,000 x g to pellet the nucleic acids; retain supernatant and discard pellet.
  - Save 200 µL of supernatant for **SAMPLE 2** = Crude Extract 2

- **Second PEI Precipitation (to pellet ClpA)**
  - Add an additional 26 µL of 10% PEI solution per mL of supernatant (creating a final concentration of 0.3% PEI).
  - Thoroughly mix for 20 min at 4°C.
  - Centrifuge the turbid mixture for 30 min at 30,000 x g to pellet ClpA; retain the pellet.
  - Save 200 µL of supernatant for **SAMPLE 3** = PEI Depleted Supernatant.

- **KCl Extraction**
  - Add 25 mL of Buffer A400 per g of original cell mass to the PEI pellet.
  - Resuspend pellet and extract soluble ClpA by stirring for 20 min at 4°C.
  - Centrifuge the KCl extract for 20 min at 30,000 x g; retain supernatant and discard pellet.
  - Save 200 µL of supernatant for **SAMPLE 4** = KCl Extract.

- **First Ammonium Sulfate Precipitation**
  - Add solid ammonium sulfate gradually to the KCl extract, to 40% saturation (0.21 g per mL extract).
  - Stir 20 minutes to overnight at 4°C
DAY 2

- **Equilibrate the HighS cation exchange column**
  - Wash the column with 5 column volumes (1 column volume = 25 mL) of 0.1 M NaOH at 1 mL/min.
  - Equilibrate the column with 5 column volumes of Buffer P100 at 6 mL/min.

- **First Ammonium Sulfate Precipitation, cont.**
  - Centrifuge the ammonium sulfate precipitate mixture (from Day 1) for 20 min at 30,000 x g; retain pellet.
  - Save 200 μL of supernatant for SAMPLE 5 = Depleted AS1 Supernatant.

- **Separate Low and High Salt Soluble ClpA**
  - Resuspend the ammonium sulfate pellet in 7 mL of Buffer P0 per g of original cell mass.
  - Check the conductivity of Buffer P100.
  - Add Buffer P0 to the resuspended pellet until the conductivity matches that of Buffer P100.
  - Mix thoroughly for 20 min at 4°C.
  - Centrifuge for 15 min at 20,000 x g; retain the supernatant.
  - Supernatant = Low Salt soluble ClpA; pellet = High Salt soluble ClpA
  - Although the pellet contains active ClpA, it has not been column purified and is therefore too unpure to use; discard pellet.
  - Save 200 μL of the supernatant for SAMPLE 6: Pre-HighS ClpA.

- **HighS Cation Exchange Purification of ClpA**
  - Add the supernatant (low salt ClpA) to the equilibrated HighS column at 1 mL/min.
  - Discard the dead volume (25 mL).
  - Collect flow through (flow through = volume of protein loaded).
  - Save 200 μL of flow through for SAMPLE 7 = HighS Flow Through.
  - Elute ClpA using a 70 mL linear gradient from P100 to P1000 at 1 mL/min; collect 2 mL (2 min) fractions (35 eluate fractions).
  - Track ClpA elution by absorbance at 280 nm and the coupled ATPase assay (see protocol below).
  - Pool active ClpA fractions.
  - Save 200 μL of pooled fractions for SAMPLE 8 = HighS Eluate.

- **Regenerate MonoS Column**
  - Wash the column with 4 column volumes of 1.0 M NaOH at 1 mL/min.
  - Wash the column with water at 6 mL/min until pH = 6-7.
  - Wash the column with 2 column volumes of 1.5 M NaCl, pH 4, at 6 mL/min
  - Wash the column with 5 column volumes of water at 6 mL/min.
  - Equilibrate the column with 20% ethanol at 6 mL/min.

- **Second Ammonium Sulfate Precipitation**
  - Add solid ammonium sulfate gradually to the HighS eluate to 70% saturation (=0.21 g per mL eluate).
  - Stir 20 min to overnight at 4°C.
DAY 3

- **Equilibrate HighQ Anion Exchange Column**
  - Wash with 1 column volume (1 column volume = 5 mL) of Buffer A100 at 2 mL/min.
  - Wash with 10 column volumes of Buffer A1000 at 6 mL/min.
  - Wash with 10 column volumes of Buffer A100 at 6 mL/min.

- **Second Ammonium Sulfate Precipitation, cont.**
  - Centrifuge ammonium sulfate precipitation mixture (from Day 2) for 20 min at 30,000 x g; retain pellet.
  - Save 200 µL of supernatant for SAMPLE 9 = Depleted AS2 Supernatant

- **Separate Low and High Salt Soluble ClpA**
  - Resuspend the ammonium sulfate pellet in 7 mL of Buffer A0 per g of original cell mass.
  - Check the conductivity of Buffer A100.
  - Add Buffer A0 to the resuspended pellet until the conductivity matches that of Buffer A100.
  - Mix thoroughly for 20 min at 4°C.
  - Centrifuge for 15 min at 20,000 x g; retain both the supernatant and the pellet.
  - Supernatant = Low Salt soluble ClpA; pellet = High Salt soluble ClpA
  - Save 200 µL of the supernatant for SAMPLE 10: Pre-HighQ ClpA.

- **HighQ Anion Exchange Purification of ClpA**
  - Add the supernatant (low salt ClpA) to the equilibrated HighQ column at 1.5 mL/min.
  - Discard the dead volume (5 mL).
  - Collect the flow through.
  - Save 200 µL of flow through for SAMPLE 11 = HighQ Flow Through
  - Elute ClpA using a 60 mL linear gradient from A100 to A1000 at 1.5 mL/min; collect 2 mL (2 min) fractions (30 eluate fractions).
  - Track ClpA elution by absorbance at 280 nm and the coupled ATPase assay (see protocol below).
  - Pool active ClpA fractions.
  - Save 200 µL of pooled fractions for SAMPLE 12 = HighQ Eluate.

- **Regenerate HighQ Column**
  - Wash the column with 4 column volumes 1.0 M NaOH at 1 mL/min.
  - Wash the column with water at 6 mL/min until pH = 6-7.
  - Wash the column with 5 column volumes of 1.5 M NaCl, pH 4, at 6 mL/min.
  - Wash the column with 10 column volumes of water at 6 mL/min.

- **High Salt ClpA: KCl Extraction**
  - Resuspend the pellet (from above) containing high salt ClpA in Buffer A400.
  - Mix thoroughly for 20 min at 4°C.
  - If the pellet does not resuspend, try adding KCl to a final concentration of 700 mM.
  - Centrifuge for 15 min at 20,000 x g; retain supernatant and discard pellet.
  - Save 200 µL of supernatant for SAMPLE 13 = High-Salt ClpA.

- **Concentrate HighQ eluate (low salt) ClpA and High Salt ClpA**
• Concentrate ClpA using Centriprep (Millipore) spin columns, MW cutoff = 30,000 Da
• Pre-wet the spin column; concentrate ClpA to > 5 mg/mL.
• Determine the concentration of purified ClpA by absorbance, $\epsilon_{280}$ for ClpA monomer ~ 0.4 mg/mL.
• Save 200 μL of low-salt ClpA for **SAMPLE 14** = Purified Low Salt ClpA.
• Save 200 μL of high-salt ClpA for **SAMPLE 15** = Purified High Salt ClpA.
• Divide purified, concentrated ClpA into aliquots of 40-80 μL.
• Flash freeze in liquid nitrogen and store at -80°C.

**For protein to be used in synchrotron footprinting studies:**
• After concentrating ClpA, perform buffer exchange into synchrotron “RXN buffer” (50mM Na Cacodylate, pH 7.0, 400 mM KCl, 20mM MgCl$_2$).
• Equilibrate a PD10 (GE Healthcare) gel filtration column with 25 mL of RXN buffer.
• Add 2.75 mL of purified, concentrated ClpA.
• Elute “buffer-exchanged” ClpA with 3.5 mL of RXN buffer.
• If ClpA looks cloudy following this step, add KCl to a final concentration of 700 mM to solubilize it.
• Save 200 μL of “buffer-exchanged” ClpA for **SAMPLE 16** = Na Cacodylate ClpA.
• Divide “buffer-exchanged” ClpA into aliquots of 40-80 μL.
• Flash freeze in liquid nitrogen and store at -80°C.

164
Below is a gel from a typical ClpA purification:

Lanes:
1 = MW markers; bands from top to bottom = 250 kDa, 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa.
2 = Crude Extract 1
3 = Crude Extract 2
4 = PEI Depleted Supernatant
5 = KCl Extract
6 = Depleted AS1 Supernatant
7 = Pre-HighS ClpA
8 = HighS Flow Through
9 = HighS Eluate
10 = Depleted AS2 Supernatant
11 = MW markers; the bands are the same as those in lane 1
12 = Pre-HighQ ClpA
13 = HighQ Flow through
14 = HighQ Eluate
15 = High Salt 2 ClpA
16 = Purified Low Salt ClpA
17 = Na Cacodylate ClpA
18 = Purified Low Salt ClpA (again, a repeat of lane 16)

*Note: As judged by gel densitometry, ClpA is ~70-75% pure following the purification protocol outlined above.
**Coupled ATPase assay used to assess activity of ClpA**

**Overview**

This ATPase assay is based on the coupling of ATP consumption by ClpA to NADH consumption by lactate dehydrogenase (LDH) via the consumption of phosphoenolpyruvate (PEP) by pyruvate kinase (PK) (see scheme below). When ClpA hydrolyzes ATP, ADP and inorganic phosphate (Pi) are produced. PK converts the ADP produced by ClpA back into ATP by the addition of a phosphate group derived from PEP, creating pyruvate. LDH then uses NADH and H+ to convert pyruvate into lactate, with concomitant formation of NAD+. NADH absorbs strongly at 340 nm; however, NAD+ does not. Therefore, the consumption of ATP by ClpA can be monitored by the reduction in NADH absorbance at 340 nm.

![Coupled ATPase assay scheme](image)

**Materials**

- **2X assay buffer**
  - 100 mM HEPES, pH 7.5
  - 600 mM KCl
  - 40 mM MgCl₂
  - 0.2% NP-40 substitute (nonylphenylpolyethylene glycol)
  - 20% glycerol

- 125 mM ATP (50X stock)
- 10 mM NADH (50X stock)
- 375 mM phosphoenolpyruvate (50X stock)
- solution of pyruvate kinase
- solution of lactate dehydrogenase
Each 500 µL assay consists of:
- 250 µL 2X assay buffer
- 10 µL ATP stock solution (final concentration = 2.5 mM)
- 10 µL NADH stock solution (final concentration = 0.2 mM)
- 10 µL phosphoenolpyruvate stock solution (final concentration = 7.5 mM)
- 9.375 units (1 unit = 1 µmol of product produced per min) pyruvate kinase
- 10.725 units lactate dehydrogenase
- ClpA sample (usually 1=10 µL)
- water to fill to 500 µL

Protocol
- Turn on the spectrophotometer and the UV lamp. Let it warm up to 37°C.
- Add everything except ClpA to a 500 µL cuvette and let it warm up (in the spectrophotometer) to 37°C (~5 min).
- Record the background absorbance at 340 nm (at 37°C) for 2 min.
- Add the ClpA sample; mix quickly by inverting the tube ~3-5 times.
- Record the absorbance at 340 nm (at 37°C) for 5 min.
- Find the slope (ΔAbs₃₄₀/Δtime) and subtract the background slope.
- Convert this slope to Δµmol of ATP/Δtime using the NADH extinction coefficient (ε₃₄₀ = 6220 M⁻¹ cm⁻¹) and the fact that 1 mol NADH consumed = 1 mol ATP consumed by ClpA

References
Appendix B: ClpP-His\textsubscript{6} purification protocol


Expression of ClpP

Overview
The plasmid encoding ClpP with a C-terminal His\textsubscript{6} tag was a generous gift from Profs. Robert Sauer and Tania Baker (MIT). The plasmid (pYK133, originating from plasmid pQE70, Qiagen) contains ampicillin resistance. This plasmid was transformed previously into *E. coli* DH5\textalpha/K175 cells; these cells contain a plasmid encoding kanamycin resistance. ClpP-His\textsubscript{6} expression was induced by the addition of isopropyl \(\beta\)-D-1-thiogalactopyranoside (IPTG).

Materials
- **LB media with kanamycin and ampicillin – 1 L**
  - 10 g tryptone
  - 5 g yeast extract
  - 5 g NaCl
  - 1 mL 1M NaOH
  - Fill to 1L with water
  - Autoclave 25 min, 121°C
  - Before starting culture add kanamycin to 30 \(\mu\)g/mL and ampicillin to 50 \(\mu\)g/mL

Protocol
- Grow 7 \(\times\) 5 mL cultures in LB media overnight
- Prepare 7 L of LB media for overexpression
- Add 5 mL of overnight starter culture per liter of LB media
- Grow at 30°C to an optical density of \(-0.6\)AU
- Induce expression of ClpP by adding IPTG to a final concentration of 0.5 mM
- Grow at 30°C for an additional 3 h
- To harvest cells, centrifuge for 15 min at 6000 x \(g\)
- Store the pellet at -80°C.
Purification of ClpP-His$_6$

Overview

ClpP-His$_6$ (referred to herein as ClpP) is purified from *E. coli* by first lysing the cells using sonication. The lysate is then loaded onto a Ni-NTA column. Proteins lacking a His-tag are washed off the column with a low concentration of imidazole (20 mM), and ClpP is then eluted in buffer containing 500 mM imidazole. This affinity-purified ClpP is dialyzed into imidazole-free buffer and purified further on a HighQ anion exchange column. This protocol typically yields 90-95% pure ClpP which can be stably stored at -80°C for several months.

Materials

- **S Buffer**: 500 mL
  - 50 mM sodium phosphate, pH 8.0
  - 1 M NaCl
  - 5 mM imidazole
  - 10% glycerol

- **W20 Buffer**: 250 mL
  - 50 mM sodium phosphate, pH 8.0
  - 1 M NaCl
  - 20 mM imidazole
  - 10% glycerol

- **W500 Buffer**: 250 mL
  - 50 mM sodium phosphate, pH 8.0
  - 1 M NaCl
  - 500 mM imidazole
  - 10% glycerol

- **Q50 Buffer**: 5 L
  - 50 mM Tris, pH 8.0
  - 10 mM MgCl$_2$
  - 50 mM KCl
  - 10% glycerol
  - 5 mM DTT (add fresh before use)

- **Q1000 Buffer**: 500 mL
  - 50 mM Tris, pH 8.0
  - 10 mM MgCl$_2$
  - 1 M KCl
  - 10% glycerol
  - 5 mM DTT (add fresh before use)

- **ClpP Buffer**: 4 L
  - 50 mM Tris, pH 7.5
  - 25 mM MgCl$_2$
  - 200 mM KCl
  - mM EDTA
  - 10% glycerol
  - 5 mM DTT (add fresh before use)
Protocol

*Note: All steps are performed at 4°C unless otherwise noted.

DAY 1

- **Lyse Cells**
  - Resuspend cell pellet in 3 mL of Buffer S per g of cells.
  - Stir to homogeneity.
  - Sonicate cells 4 x 1 min (2 s on, 1 s off; 70% amplitude).
  - Centrifuge for 20 min at 17,000 x g to clear the lysate; retain supernatant and discard pellet.
  - Save 200 µL of supernatant for SAMPLE 1 = Crude Extract.

- **Equilibrate Ni-NTA Resin**
  - To create a column volume of 7.5 mL Ni-NTA resin, use 15 mL of a 50% resin slurry.
  - In a 50 mL conical, wash 5 times with 15 mL Buffer S; centrifuge at low speed (1000 rpm on tabletop centrifuge, Beckman S4180 rotor) for 5 min between washes.
  - Can do this step the night before (can store in Buffer S for 12 h or less)

- **Load Ni-NTA Resin and Pour Column**
  - Add cleared lysate to equilibrated Ni-NTA resin.
  - Mix 1 h at 4°C on a shaker.
  - Pack resin into column; collect flow through (flow through= volume of lysate loaded)
  - Save 200 µL of flow through for SAMPLE 2 = Ni-NTA Flow Through.
  - Insert flow controller (Bio-Rad) to bed level; tighten flow controller in place and tighten O-ring; snap closed.

- **Ni-NTA Chromatography**
  - Wash the column with 200 mL of Buffer S at 1.6 mL/min.
  - Wash the column with 100 mL of Buffer W20 at 1.6 mL/min. Collect 5 min fractions (~12 tubes).
  - Elute ClpP with 51.2 mL of Buffer W500 at 1.6 mL/min. Collect 1 min fractions (~32 tubes).
  - Track ClpP elution by absorbance at 280 nm and the SLY-AMC activity assay (see protocol below).
  - Pool active ClpP fractions
  - Save 200 µL of pooled fractions for SAMPLE 3 = Ni-NTA Eluate.

- **Dialyze Affinity-Purified ClpP**
  - Prepare dialysis buffer: add 3.1 g of DTT to 4 L Buffer Q50 for a final concentration of 5 mM DTT.
  - Wash dialysis tubing (molecular weight cutoff ~3-5 kDa) with water followed by buffer Q50 (15-30 min soak in each).
  - Dialyze Ni-NTA eluate in 4L Buffer Q50 at 4°C overnight.
- **Regenerate Ni-NTA Column**
  - Wash the column with 5 column volumes (1 column volume = 7.5 mL) of 0.5 M NaOH over 30 min.
  - Wash the column with 5 column volumes of 30% ethanol.
  - Store the column in 30% ethanol.

**DAY 2**

- **Equilibrate HighQ Anion Exchange Column (column volume = 5 mL)**
  - Prepare column buffers: add 300 µl of 1M DTT to 300 mL each of Buffers Q50 and Q1000.
  - Wash the column with 4 mL of Buffer Q50 (2 mL/min x 2 min).
  - Wash the column with 60 mL of Buffer Q1000 (6 mL/min x 10 min).
  - Equilibrate the column with 60 mL of Buffer Q50 (6 mL/min x 10 min).

- **Load HighQ Anion Exchange Column**
  - Filter dialyzed protein through 0.45 µm filter.
  - Save 200 µl of filtered protein for **SAMPLE 4** = Dialyzed Ni-NTA Eluate.
  - Add filtered, dialyzed protein to the equilibrated column at 1.5 mL/min.
  - Discard the dead volume (5 mL). Collect the flow through.
  - Save 200 µl of the flow through for **SAMPLE 5** = HighQ Flow Through.

- **HighQ Anion Exchange Column Chromatography**
  - Wash the column with 2 column volumes of Buffer Q50.
  - Elute ClpP with a 50 mL linear salt gradient from Buffer Q50 to Buffer Q1000 at 1.5 mL/min; collect 1.5 min (2.25 mL) fractions (~22 fractions).
  - Track ClpP elution by absorbance at 280 nm and the SLY-AMC activity assay (see protocol below).
  - Pool active ClpP fractions.
  - Save 200 µl of pooled fractions for **SAMPLE 6** = HighQ Eluate.

- **Dialyze Purified ClpP**
  - Add 0.62 g DTT to 4L ClpP Buffer for a final concentration of 1 mM DTT.
  - Wash dialysis tubing with water followed by ClpP Buffer (15-30 min soak in each).
  - Dialyze HighQ-purified ClpP in 4L ClpP Buffer at 4°C overnight.
  - Save 200 µl of dialyzed ClpP for **SAMPLE 7** = Dialyzed, Purified ClpP.
  - Divide purified, dialyzed ClpP into aliquots of 40-80 uL.
  - Flash freeze in liquid nitrogen and store at -80°C.

- **Regenerate HighQ Anion Exchange Column**
  - Wash with 30 mL of 1 M NaOH at 1 mL/min.
  - Wash with 50 mL of water at 6 mL/min.
  - Wash with 25 mL of Q1000 at 6 mL/min.
  - Wash with 50 mL of water at 6 mL/min.
- Wash with 50 mL of 20% ethanol at 6 mL/min; store in 20% ethanol.

**Buffer Exchange (for protein to be used in synchrotron footprinting studies):**
- Equilibrate a PD10 (GE Healthcare) gel filtration column with 25 mL of synchrotron “RXN buffer” (50mM Na Cacodylate, pH 7.0, 400 mM KCl, 20mM MgCl₂).
- Add 2.75 mL of purified ClpP.
- Elute “buffer-exchanged” ClpP with 3.5 mL of RXN buffer.
- Save 200 μL of “buffer-exchanged” ClpP for SAMPLE 8 = Na Cacodylate ClpP.
- Divide “buffer-exchanged” ClpP into aliquots of 40-80 μL.
- Flash freeze in liquid nitrogen and store at -80°C.

**Below is a gel from a typical ClpP purification:**

![Gel Image]

**Lanes:**
1 = MW markers; bands from top to bottom = 250 kDa, 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa, 15 kDa, 10 kDa
2 = Crude Extract
3 = Ni-NTA Flow Through
4 = Ni-NTA Eluate
5 = Dialyzed Ni-NTA Eluate
6 = HighQ Flow Through
7 = HighQ Eluate
8 = Na Cacodylate ClpP
9 = Dialyzed, Purified ClpP

*Note: As judged by gel densitometry, ClpP is ~90-95% pure following the purification protocol outlined above.
Succ-Leu-Tyr-AMC (SLY-AMC) activity assay

Overview
This assay assesses the proteolytic activity of ClpP using a model substrate, succinyl-leucine-tyrosine-7-amido-4-methylcoumarin (Succ-Leu-Tyr-AMC). Succ-Leu-Tyr-AMC has low fluorescence, however, cleavage between the tyrosine and AMC moieties by ClpP releases 7-amino-4-methylcoumarin (AMC), which is highly fluorescent (ex: 345 nm, em: 440 nm, see scheme below). Therefore, the proteolytic activity of ClpP can be measured by monitoring the increase in AMC fluorescence over time.

Materials
- 10X assay buffer
  - 500 mM Tris-HCl, pH 8.0
  - 1 M KCl
  - 10 mM DTT (add fresh before use)
- 20 mM Succ-Leu-Tyr-AMC in DMSO (20X stock)
- Each 100 μL assay consists of:
  - 10 μL 10X assay buffer
  - 5 μL Succ-Leu-Tyr-AMC stock solution (final concentration = 1 mM)
  - ClpP sample (~0.5 μg, usually 1-5 μL)
  - water to fill to 100 μL
Protocol

- Turn on the fluorimeter, and let it warm up to 37°C.
- Add everything except ClpP to one well of a 96-well plate. Let this warm up (in the fluorimeter) to 37°C (~5 min).
- Add the ClpP sample.
- Record the fluorescence (ex: 345 nm, em: 440 nm) for 5-10 min (at 37°C).
- Find the slope ($\Delta$RFU/$\Delta$time)
- Convert this slope to $\Delta \mu$mol AMC/$\Delta$time using an AMC standard curve.
Appendix C: GFP-ssrA purification protocol


**Expression of GFP-ssrA**

**Overview**
A plasmid expressing GFP containing the S65G and S72A mutations, which enhance the intensity of green fluorescence [1], and a C-terminal ssrA tag was kindly provided by Prof. Sren Molin (BioCentrum DTU, Denmark). This plasmid contains ampicillin resistance and was transformed previously into *E. coli* JB401 cells. GFP-ssrA expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG).

**Materials**
- LB media with ampicillin- 1L
  - 10 g tryptone
  - 5 g yeast extract
  - 5 g NaCl
  - 1 mL 1M NaOH
  - Fill to 1L with water
  - Autoclave 25 min, 121°C
  - Before starting culture add ampicillin to 50 μg/mL

**Protocol**
- Grow 4 x 5 mL cultures in LB media overnight
- Prepare 4L of LB media for overexpression
- Add 5 mL of overnight starter culture per L of LB media
- Grow at 37°C to an optical density of ~0.6 AU
- Induce expression of GFP-ssrA by adding IPTG to a final concentration of 250 μM
- Grow at 25°C for an additional 6 h
- To harvest cells, centrifuge for 15 min at 6000 x g
- Store the pellet at -80°C overnight or for up to a month
Purification of GFP-ssrA

Overview

GFP-ssrA is purified from *E. coli* by first lysing the cells using sonication. Other proteins are precipitated from the lysate by the addition of ammonium sulfate (to 40% saturation) and triethanolamine (to 100 mM). GFP-ssrA is extracted from the supernatant through a series of organic extractions. Finally, extracted GFP-ssrA purified further on a phenyl sepharose column. This protocol typically yields 90-95% pure GFP-ssrA which can be stably stored for several months in the dark at 4°C.

Materials

- **Buffer A**
  - 20 mM Tris-HCl, pH 8.0
  - 150 mM NaCl
  - 5 mM EDTA

- **Buffer B**
  - 20 mM Tris-HCl, pH 8.0
  - 1 mM EDTA
Protocol

*Note: All steps are performed at room temperature unless otherwise noted.

- **Lyse cells (4°C)**
  - Resuspend cell pellet in 4 mL of Buffer A per g of cells.
  - Sonicate cells for 15 min (1 s on / 4 s off; 70% amplitude)
  - Centrifuge for 15 min at 30,000 x g to clear lysate; retain supernatant and discard pellet.
  - Save 200 μL of supernatant for **SAMPLE 1** = Crude Extract.

- **Ammonium Sulfate / TEA Precipitation (to remove other proteins)**
  - Add ammonium sulfate to 40% saturation (0.21 g per mL of lysate)
  - Add triethanolamine (TEA) to 100 mM (13.32 μL per mL of lysate)
  - Stir for 1 h
  - Centrifuge for 20 min at 30,000 x g; discard pellet and retain supernatant.
  - Save 200 μL of supernatant for **SAMPLE 2** = Crude GFP1.

- **First Ethanol Extraction**
  - Add ammonium sulfate to 70% saturation (0.3675 g per mL crude GFP1)
  - Add 1/4 volume ethanol to salted crude GFP
  - Shake vigorously for 1 min.
  - Centrifuge for 5 min at 3,000 x g to separate phases.
  - Collect GFP in the upper organic phase.

- **Second Ethanol Extraction**
  - Add 1/16 volume ethanol to aqueous phase from first ethanol extraction.
  - Shake vigorously for 1 min.
  - Centrifuge for 5 min at 3,000 x g to separate phases.
  - Collect GFP in the upper organic phase; combine with previous GFP organic phase.
  - Save 200 μL of combined organic phase for **SAMPLE 3** = Ethanol GFP Extract.

- **n-Butanol, chloroform, and ammonium sulfate extractions**
  - Add 1/4 volume n-butanol to ethanol GFP extract.
  - Shake vigorously for 30 s.
  - Centrifuge for 5 min at 3,000 x g to separate phases; discard upper phase.
  - Repeat with a smaller volume of n-butanol.
  - To the lower phase:
    - Add 1 volume chloroform to lower aqueous phase
    - Shake vigorously for 30 s.
    - Centrifuge for 5 min at 3,000 x g to separate phases.
    - Save upper aqueous phase.
  - To the lower phase:
    - Add 1 volume of 30% saturated ammonium sulfate in water.
    - Shake vigorously for 30 s.
    - Centrifuge for 5 min at 3,000 x g to separate phases.
    - Combine resulting aqueous phase with previous aqueous phase.
  - Filter combined protein extract through 0.45 μm filter.
• Save 200 ul of combined protein extract for **SAMPLE 4 = Salt GFP Extract.**

  - **Phenyl Sepharose Chromatography**
    - Wash the column with 5 column volumes (1 column volume = 25 mL) of 20% ammonium sulfate in Buffer B at 3 mL/min.
    - Equilibrate the column with 20% ammonium sulfate in Buffer B for 3 mL/min for 40 min.
    - Load Salt GFP Extract onto column at 1.5 mL/min.
    - Discard dead volume (25 mL).
    - Collect flow through (= volume of Salt GFP Extract loaded onto column).
    - Save 200 µl of flow through for **SAMPLE 5 = Phenyl Sepharose Flow Through.**
    - Elute GFP with a 40 mL linear gradient from 20% to 0% ammonium sulfate in Buffer B at 1 mL/min.
    - Collect 2 mL (2 min) fractions (~20 fractions total).
    - Take GFP elution by absorbance at 280 nm and fluorescence ($\lambda_{ex} = 467$ nm; $\lambda_{em} = 512$ nm); combine GFP-containing fractions.
    - Save 200 µl of combined fractions for **Sample 6 = pure GFP.**
    - Store purified GFP at or below 4°C in the dark.

  - **Regenerate column**
    - Wash the column with 80 mL 1 M NaOH at 1 mL/min.
    - Wash the column with 60 mL water at 6 mL/min.
    - Equilibrate the column with 50 mL of 20% ethanol at 6 mL/min.
    - Store the column in 20% ethanol.

**References**

LAURA DANIELLE JENNINGS

DEPARTMENT OF CHEMISTRY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
32 VASSAR ST. BLDG 56-522
CAMBRIDGE, MA 02139
617.257.0106 (CELL)
EMAIL: LAURAJ@MIT.EDU

EDUCATION

<table>
<thead>
<tr>
<th>Year</th>
<th>Institution</th>
<th>Location</th>
<th>Degree</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-2008</td>
<td>Massachusetts Institute of Technology</td>
<td>Cambridge, MA</td>
<td>Ph.D., Biological Chemistry</td>
<td>5.00/5.00</td>
</tr>
<tr>
<td>1999-2004</td>
<td>University of Missouri</td>
<td>Columbia, MO</td>
<td>B.S., Chemistry and Mathematics</td>
<td>4.00/4.00</td>
</tr>
</tbody>
</table>

RESEARCH EXPERIENCE

<table>
<thead>
<tr>
<th>Year</th>
<th>Institution</th>
<th>Location</th>
<th>Position</th>
<th>Advisor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-2008</td>
<td>Massachusetts Institute of Technology</td>
<td>Cambridge, MA</td>
<td>Graduate Research Assistant, Department of Chemistry</td>
<td>Prof. Stuart Licht</td>
</tr>
<tr>
<td>Summer 2003</td>
<td>Mayo Clinic Graduate School</td>
<td>Rochester, MN</td>
<td>Summer Undergraduate Research Fellow</td>
<td>Prof. Matthew Ames, Department of Pharmacology</td>
</tr>
<tr>
<td>Fall 2002</td>
<td>University of Missouri</td>
<td>Columbia, MO</td>
<td>Research Experience for Undergraduates Program</td>
<td>Prof. Peter Tipton, Department of Biochemistry</td>
</tr>
<tr>
<td>Summer 2002</td>
<td>Texas A&amp;M University</td>
<td>College Station, TX</td>
<td>NSF Research Experience for Undergraduates Program</td>
<td>Prof. Paul Fitzpatrick, Department of Biochemistry</td>
</tr>
</tbody>
</table>

PUBLICATIONS


L.D. Jennings, J. Bohon, M.R. Chance, S. Licht. The ClpP N-terminus coordinates substrate access with protease active site reactivity; accepted for publication in *Biochemistry*.

L.D. Jennings, D.S. Lun, M. Médard, S. Licht. ClpP hydrolyzes a protein substrate processively and independently of the ClpA ATPase: Mechanistic studies of ATP-independent processive proteolysis; accepted for publication in *Biochemistry*.
### PRESENTATIONS

<table>
<thead>
<tr>
<th>Year</th>
<th>Event, Institution, Location, Title</th>
</tr>
</thead>
</table>

### TEACHING EXPERIENCE

<table>
<thead>
<tr>
<th>Year</th>
<th>Institution, Location</th>
<th>Position, Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring 2005</td>
<td>Massachusetts Institute of Technology, Cambridge, MA</td>
<td>Teaching Assistant, Biochemistry Laboratory</td>
</tr>
<tr>
<td>Fall 2004</td>
<td>Massachusetts Institute of Technology, Cambridge, MA</td>
<td>Teaching Assistant, Introductory Biochemistry</td>
</tr>
<tr>
<td>2002-2004</td>
<td>University of Missouri, Columbia, MO</td>
<td>Teaching Assistant, General Chemistry</td>
</tr>
<tr>
<td>Summer 2004</td>
<td>KAPLAN Test Prep and Admissions, Columbia, MO</td>
<td>MCAT Physics Instructor</td>
</tr>
<tr>
<td>2001-2002</td>
<td>University of Missouri, Columbia, MO</td>
<td>Student Success Center, Mathematics and Chemistry Tutor</td>
</tr>
</tbody>
</table>

### SELECTED ACADEMIC HONORS

- Recipient of Graduate and Postdoctoral Travel Award for 2007 ASBMB annual meeting
- 2004-2005 Outstanding Teaching Assistant Award, MIT
- Phi Beta Kappa National Honor Society Inductee, 2004
- Barry M. Goldwater Scholar 2003
- Mortar Board National Honor Society Inductee, 2002
- Student Spotlight Feature in @Mizzou Online Alumni Newsletter, 2003
- Phi Lambda Upsilon National Chemistry Honor Society, 2002-2004
- CRC Freshman Outstanding Chemistry Student of the Year, Univ. of Missouri, 2000
- Phi Eta Sigma National Honor Society, 2000-2004
- University of Missouri College of Arts and Sciences Award of Excellence Scholar 2002-2003
- Missouri Top 100 Scholar, 1999
  awarded yearly to the top 100 high school graduates in the state of Missouri

### SELECTED LEADERSHIP, VOLUNTEER, AND EXTRACURRICULAR EXPERIENCE

- MIT Women in Chemistry (WIC) Administrative Committee, 2005-2006
- MIT Chemistry Graduate Student Committee member, 2004-present, President, 2006-2007
- MIT Chemistry Outreach Member, 2005-2006
- Country Music 1/2 Marathon Participant, Nashville, TN, 2003 and 2004
- University of Missouri Homecoming Steering Committee, 2002
- Study Abroad Program, Imperial College, London, England, Summer 2001
- Kappa Alpha Theta National Panhellenic Sorority, 1999-2003, Officer, 2001
- RAMS (Rockin’ Against Multiple Sclerosis) Steering Committee, 2001 and 2002